VOL. 10 No. 2, pp. 37-68 AUGUST 1948

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS.

A DIAGNOSTIC X-RAY APPARATUS WITH EXPOSURE TECHNIQUE INDICATION AND OVERLOAD PROTECTION

By A. NEMET *), W. A. BAYFIELD *), and M. BERENDEI *).

621.386.1:
616-073.75

In order to obtain radiographs of the highest possible quality with a given diagnostic X-ray apparatus, in general it is desirable to choose the voltage at the X-ray tube, the tube current and the exposure time in such a way that the tube is loaded to the permissible focus temperature. With a universal diagnostic apparatus described earlier in this Review, this was achieved by allowing the operator to select the voltage and the exposure time only, the current then being adjusted automatically to its maximum permissible value. In many cases, however, a certain freedom of choice, together with a clear insight of other possible settings will be preferred. Owing to these considerations, Philips Electrical Ltd, London, has developed an apparatus featuring a clear survey of all permissible and non-permissible settings by means of a special signalling system, and an overload protection which on selecting a non-permissible setting (i.e. which would cause the X-ray tube to be overloaded) automatically renders impossible the switching on of the high tension generator. For complete reliability of the protection and the signalling system, severe demands were to be imposed on the reproducibility of the voltage, current and time setting. Among the measures taken with a view to meeting these requirements and described in this article, one of the most interesting is that regarding the voltage drop in the tube circuit: this voltage drop for every mA setting is brought to a constant value by means of a bank of resistances, from which the appropriate one is selected by the mA control handle and connected in series with the primary coil of the high tension transformer.

With fixed tube voltage, the radiation output of an X-ray tube increases proportionally as the tube current is increased. A limit to this is imposed by the development of heat in the focal spot: the temperature of the anode at this spot must not exceed a certain value. The current rating with which the permissible temperature will be obtained depends on the time during which the tube current flows.

Thus, when making a radiograph with a given X-ray tube, for every choice of tube voltage and exposure time a maximum permissible value of the tube current will exist. The relation between these three variables may be represented by a diagram with a series of curves, the so-called rating chart. As an example, *fig. 1* shows the rating chart of the "Rotalix" X-ray tube with 2 mm focus when working on full wave rectification.

With a given tube, where there is involuntary movement, sharper images will (other things being equal) be obtained on increasing the tube current, as this diminishes the exposure time necessary for producing the desired film density. Therefore, in general, it is desirable that the operator, when choosing tube voltage, current and exposure time for a radiograph, goes right to the limit of what he is allowed by the rating chart.

The designer of X-ray apparatus can proceed in different ways to aid the operator in using the tube always in the described optimal manner. The most radical solution of the problem consists in abolishing entirely the free choice of the tube current by the operator and in ensuring instead, by means of a suitable control mechanism, that on choosing the voltage and the exposure time the current is rated automatically to its maximum permissible value. A diagnostic apparatus with this principle incorporated was described several years ago in this Review ¹).

---

*) X-Ray Development Laboratory, Philips Electrical Ltd, London.

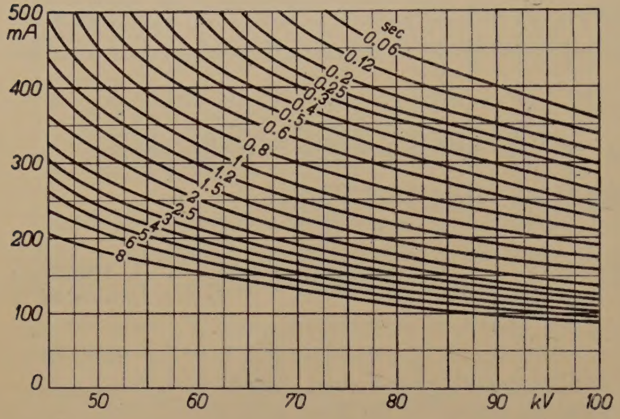

Fig. 1. Rating chart of a "Rotalix" X-ray tube with 2 mm focus when loaded with full wave rectified A.C. For a number of values of the exposure time, taken as a parameter, the maximum permissible tube current is plotted against the adjusted voltage (in kV$_{\mathrm{peak}}$). Thus, for every combination of tube voltage, tube current and exposure time obtained from the chart, the focus is heated exactly to the permissible temperature.

There is no doubt that this method has important advantages and gives the most satisfactory results. There remains, however, the possibility that some operators may find that complete automatism is a disadvantage under certain circumstances. For instance, where it is desired to follow a known exposure technique or to develop and reproduce new exposure factors, greater flexibility is a desirable feature. Furthermore, with the completely automatic method referred to above it is not practicable for the manufacturer to make an apparatus delivering for every one of the hundreds of combinations of possible voltage and time adjustments the exact value of the corresponding permissible current. In a practical approximation, the tube current may be underrated by say 30 % for the most unfavourable combination ²). Although in normal cases this does not mean an appreciable loss of quality of the radiograph, it is quite likely that for difficult objects the operator will miss this difference.

So it may be stated that, in addition to the radical method described, which is based on a complete automatisation, there is room for a less radical method, asking more deliberation from the operator and on the other hand offering him more latitude and more possibilities. A diagnostic apparatus (DX 4) making use of such a method and developed by Philips Electrical Ltd, London, will be discussed in this paper.

General description of the apparatus

As the apparatus, apart from the control method and some other features which will be described in detail, is designed on the same general lines as the diagnostic apparatus previously reported in this Review ¹), we shall only give a brief general description.

The apparatus consists of two units: the high tension generator cabinet — housing the high tension generator with its auxiliary gear and a remote controlled high tension switch — and the control table; these are linked by a flexible connecting cable. A maximum of three different X-ray tubes, one of which may be of the double focus type ³), may be connected to the generator unit, so that

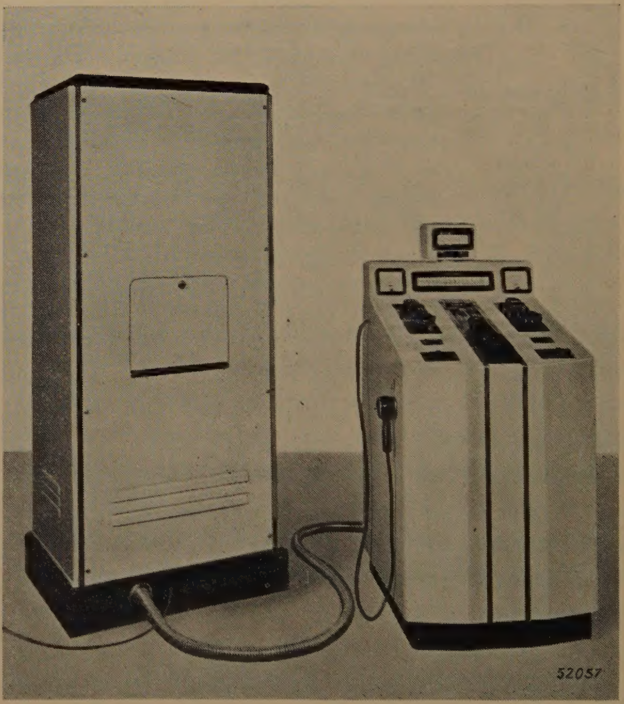

Fig. 2. The universal diagnostic X-ray apparatus described in this paper (D X 4) consists of two units: on the left hand the high tension generator cabinet, containing the high tension transformer, rectifying valves and a high tension switch, for supplying various X-ray tubes; on the right hand the control table containing the switch gear for energizing the X-ray tube for taking radiographs and for screening, control apparatus for making the desired settings for a radiograph, various meters and auxiliaries. A hand switch for taking exposures is attached to the control table by a flexible lead.

---

¹) H. A. G. Hazeu & J. M. Ledeboer, A universal diagnostic X-ray apparatus, Philips Techn. Review **6**, 12-21, 1941.

²) Cf. the article mentioned under ¹).

³) A tube of this type is dealt with e.g. in Philips Techn. Review **8**, p. 327 (fig. 7), 1946 (No. 11).

four focal spots are available. By means of a focus selector switch in the control desk and the high tension switch mentioned above, any one of these may be selected.

The generator contains a high tension transformer and four rectifying valves of the high vacuum type in a bridge circuit. Thus, the output of the generator is pulsating D.C. By means of an autotransformer, the primary voltage of the high tension transformer is controlled so that the tube voltage may be varied between 45 and 100 kV_{peak} in 23 steps of $2^{1}/_{2}$ kV.

The generator also contains two filament transformers, for feeding the filaments of double focus X-ray tubes. A variable resistance in the primary circuit of these transformers permits the adjustment of the output current in 10 steps from 25 to 500 mA, mean value.

The switching of the high tension for producing a radiograph is effected by a time switch operating a high speed contactor, allowing 26 different time settings, from 0.02 to 9 seconds. During screening the time switch is inoperative, the tube current in this case being adjustable from 1 to 5 mA by means of a stepless control. The tube voltage control for screening is also stepless.

All these adjustments are read on calibrated scales on the control table.

In addition to the switches mentioned, the control table also contains a number of auxiliary switches such as room light switches, fan switches for X-ray tubes with fan cooling and a selector switch for the selection of external accessories such as a Potter-Bucky diaphragm, a serial changer for stomach radiography, a planigraph movement of tube and film, and so on. The limited scope of this article does not permit the description of all these more or less conventional details. As a substitute we give a few photographs of the complete apparatus and of the control table (*figs. 2* to *4*), the legends conveying some further information. Several important features will be discussed in connection with the control method, which we shall describe in the following section.

Control method and automatic overload protection of the tube

The idea of the control method incorporated in the present apparatus [4]) is to give the operator, when selecting voltage, current and exposure time, at any moment a clear survey of the relation between the combination actually selected and the series of other possible combinations, so that he may see how far he has approached the limit. At the same time the apparatus must contain a safeguard automatically rendering the exposure impossible when the operator has selected a combination which is not permissible, i.e. which would cause overloading of the tube.

Overload protection

We shall first consider the realization of the second object, the protection of the tube. For this purpose a relay contact is connected in series with the energizing circuit of the main contactor, the

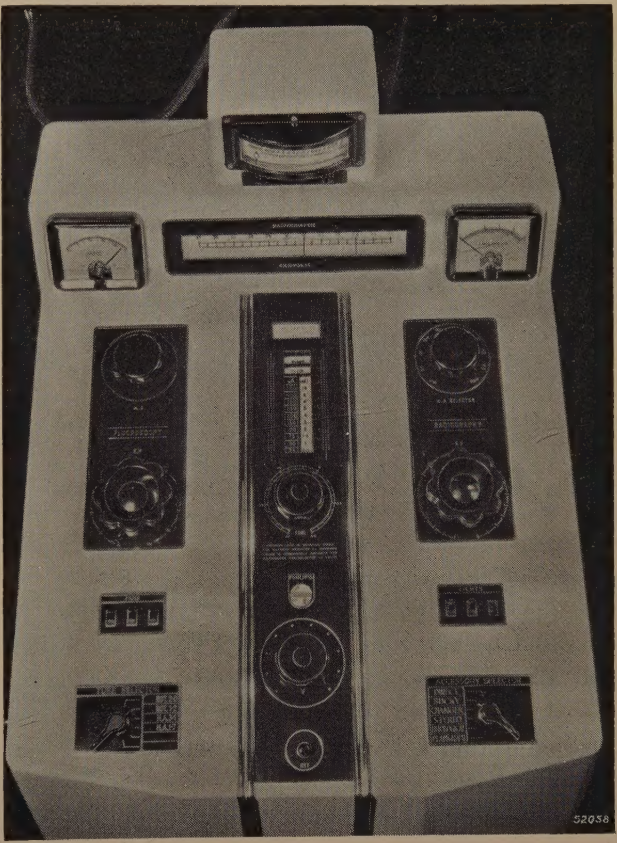

Fig. 3. Desk of the control table. On the left: the two control knobs for the tube current and tube voltage on screening; on the right: the corresponding handles for taking radiographs; in the centre; the exposure time selector with the signalling system described below. In the lower part: the tube and focus selector handle (left), a selector switch for various external accessories (right) and between these the "off" switch of the apparatus and the mains compensating switch. The mains compensation is checked on the meter on top left part, the meter on the opposite side indicating the tube current on long exposures (1—8 secs.). The same meter indicates on a separate scale the tube current on screening, and by a simple change-over it can be used for reading the filament current of the X-ray tube. This is an important help, as small deviations of the filament current, due to errors in the pre-setting or to small variations in the resistances of contacts, would impair the control technique described below. The top centre meter indicates the mAsec: it is switched on automatically with exposures of less than 1 sec; beneath this meter is the indicator of tube kilovoltage for radiography. The kilovoltage for screening is engraved on the corresponding control knob.

---

[4]) A. F. Jeans, Brit. Pat. 571243.

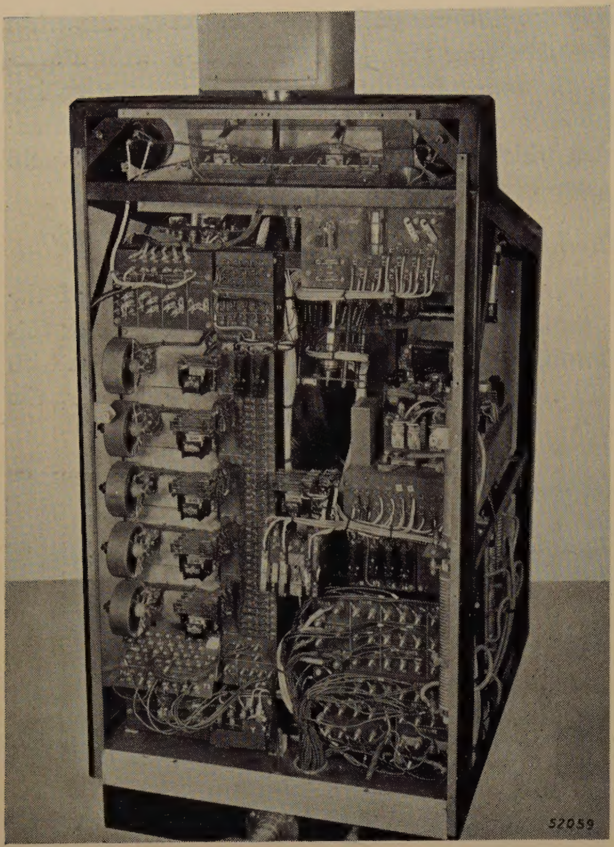

Fig. 4. Back view of the control table with panel removed. To the right from the centre the time switch is visible, to the left the control units for the setting of the tube current for the four different foci. The control mechanism proper is not visible on this photograph.

relay coil being supplied by the current in an auxiliary circuit. Whether the contactor can be operated by the time switch and the tube loaded will thus depend on the auxiliary circuit being closed. This circuit is interrupted in one place, and the connection between the two terminals, P_1 and P_2, is controlled by the selection of the tube voltage, tube current and exposure time in the following manner.

By means of selector arms (A_1, A_2) mounted on the shafts of the control knobs for the tube voltage and tube current, one of the terminals (P_1 in *fig. 5*) is connected, for every combination of selected voltage and current, to a corresponding contact stud S_i. The number of studs S provided is equal to the number of possible combinations of kV values and mA-values, i.e. in our case 23·10 = 230. Let us imagine that these contact studs are arranged on a rectangle, all studs corresponding to a common kV value lying on a vertical line and all studs with a common mA value on a horizontal line. Hence, the studs represent a picture of the coordinate system of the rating chart; cf. Fig. 5.

The second terminal (P_2) of the auxiliary circuit ends in a selector arm coupled to the exposure time control knob, which for every time setting connects this terminal to a corresponding contact stud of a series of 26 studs T.

Each contact stud S_i is linked permanently to the contact stud T_k belonging to the longest exposure time permissible with the combination "S_i" of voltage and current. Thus, when this "T_k" with the voltage-current combination "S_i" are selected, the auxiliary circuit is closed and the exposure can take place.
There exist, however, a number of other voltage-current combinations with the same maximum permissible exposure time "T_k". Such combinations are indicated by the curve of the rating chart belonging to that time (Fig. 1). All contact studs S corresponding to these combinations are also permanently connected to the contact T_k and hence linked mutually. These chains of connections, each chain belonging to one of the contacts T,

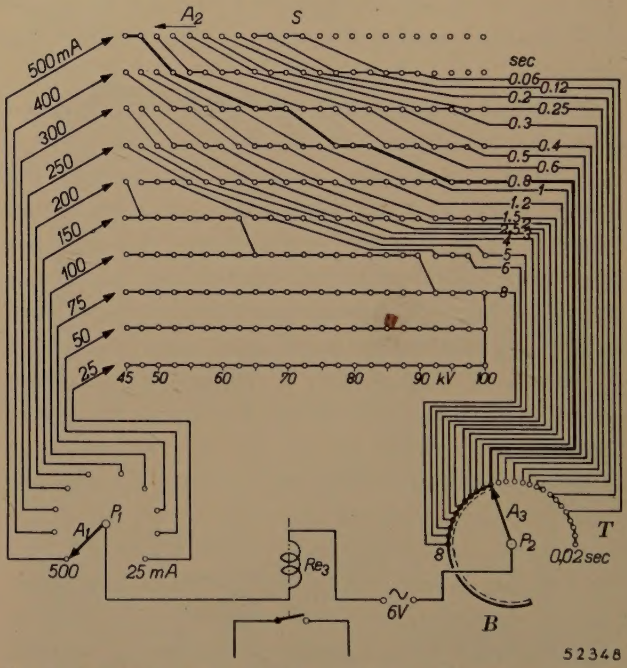

Fig. 5. Schematic representation of the auxiliary circuit used for the overload protection of the X-ray tube. If relay Re_3 is de-energized, the supply for the high tension transformer is interrupted and no exposure is possible. This depends, therefore, on a connection being established or not between the terminals P_1 and P_2. By means of selector arms A_1, A_2 coupled to the control handles for the current and the voltage settings (the form of A_2 will be explained in the following), the terminal P_1 is connected to one of the 230 contact studs S corresponding to the 230 possible voltage-curernt combinations. The selector arm A_3, coupled to the time control knob, connects terminal P_2 to one of the 26 contact studs T belonging to the 26 possible exposure time values. B is a wiper arm fastened to A_3. Due to the connections traced between chains of contact studs S on the one hand and contact studs T on the other hand, overload protection is achieved, as is explained in the article. (The connections traced apply to a "Rotalix" X-ray tube with chart shown in Fig. 1.)

appear as the series of chart curves in the coordinate system of the studs S in Fig. 5.

Finally, the selector arm of the contacts T is fitted with a wiper arm shown in the figure as a circular segment, which on selecting a contact stud T_k shortcircuits this stud with all those belonging to exposure times longer than T_k. Hence, when selecting T_k and then taking a voltage current combination for which a still longer exposure time than T_k would be permissible, the auxiliary circuit again will be closed and the overload protecting relay will be energized.

When, on the other hand, a time T_k is selected together with a voltage-current combination which would only allow a shorter exposure time than T_k, the auxiliary circuit remains open, as can be easily verified in Fig. 5; the relay is de-energized and the main contactor remains inoperative. In this way the overload protection of the tube is achieved.

As to the actual design of the system described, only the selection of the contact stud S_i needs some further consideration. The rectangle containing 23 columns of studs S and represented in fig. 5 is in fact bent round a circular drum coupled to the kV selector switch. A vertical row of ten brushes is in contact with this drum (at A_2). Depending on the position of the drum, the brushes are in contact with studs S of one of the 23 columns, this column corresponding to the kV value selected. The previously mentioned selector arm of the mA-control knob (A_1) connects the terminal P_1 of the auxiliary circuit to one of the ten brushes, each brush belonging to a fixed mA value. Thus, the terminal P_1 is connected with the contact stud S_i belonging to the selected voltage-current combination. The permanent links between the studs T on the one hand, and the chains of mutually connected studs S on the drum on the other hand, are maintained by means of series of slip rings at the end of the drum.

Fig. 6 shows a photograph of the drum with contact studs S. For every X-ray focus which may be fed through the apparatus a different system of mutual connections of studs S is required, corresponding to the rating chart of the focus under consideration. Therefore the drum contains four coordinate systems of studs as pictured in Fig. 5, mounted one above the other. The drum section belonging to the selected focus is brought into action by the focus selector switch on the control table. As generally only one of two foci allowing the maximum mA values of 400 or 500 mA will be used, two of the sections are fitted with seven rows of studs only instead of ten as described above. The wiring of the contact studs and slip rings is placed inside the drum.

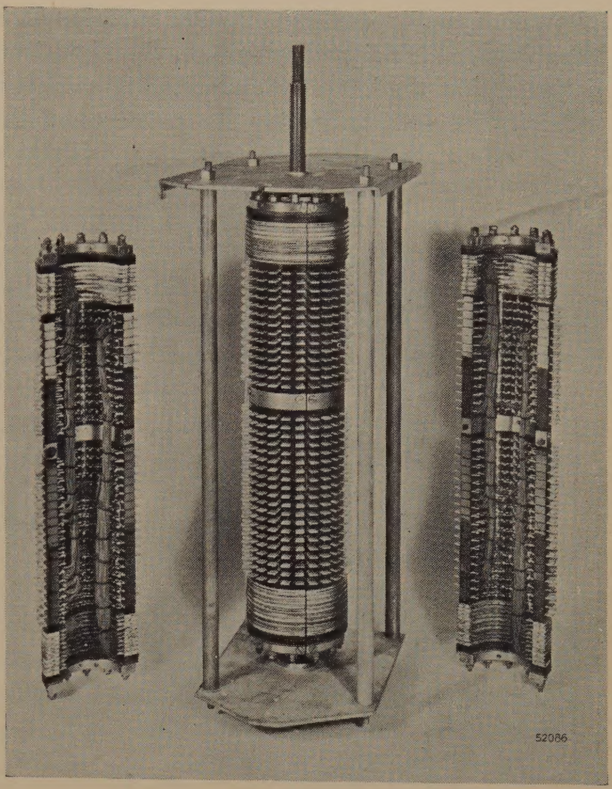

Fig. 6. Control drum with the 230 contact studs S from Fig. 5. The drum is mounted on the shaft of the kV control handle and contains in four sections four assemblies of similar contact studs S, corresponding to the four different foci to be energized by the apparatus. The wiring between studs S is placed within the drum, as is to be seen in the two half drums shown separately. The permanent connections with studs T (cf. Fig. 5) are maintained by means of slip rings visible on the ends of the drum (D in Fig. 7).

Signalling

The second object stated in the beginning of this chapter, viz. to give the operator for every setting a survey of the possible variations, is achieved by means of a signalling system making use of the auxiliary circuit already introduced. To each of the ten brushes mentioned above and corresponding to the ten possible mA values, a lamp is connected in the manner indicated in *Fig. 7*. The ten lamps are mounted on the control table under a window. With the aid of Fig. 7 it may be seeen that on selecting a kV value and a time setting all lamps will be alight that belong to mA values permissible for this combination (regardless of the actual mA setting as yet to be chosen). This enables the operator to see at a glance the range of mA settings from which he can choose.

For practical purposes generally not the mA value is regarded as of prime importance, but the number of mAsec, this number determining the photographic density obtained with a fixed tube voltage. The signalling system, therefore, is so designed that the lamps when lighted illuminate a column of figures indicating the products of the mA values with the selected exposure time. This is obtained by arranging between the lamps and the window in the control desk a transparent rotating disc, coupled to the time selector handle and carrying 26 radial columns of ten mAsec figures each. Depending on the time setting, a different series of mAsec figures is visible through the window. Moreover, the window shows another column containing the ten (fixed) mA values, of which only the value adjusted at that moment is illuminated (by a simple signalling circuit not indicated in Fig. 7). This second column is placed opposite to the mAsec column, as is to be seen in *Fig. 8*.

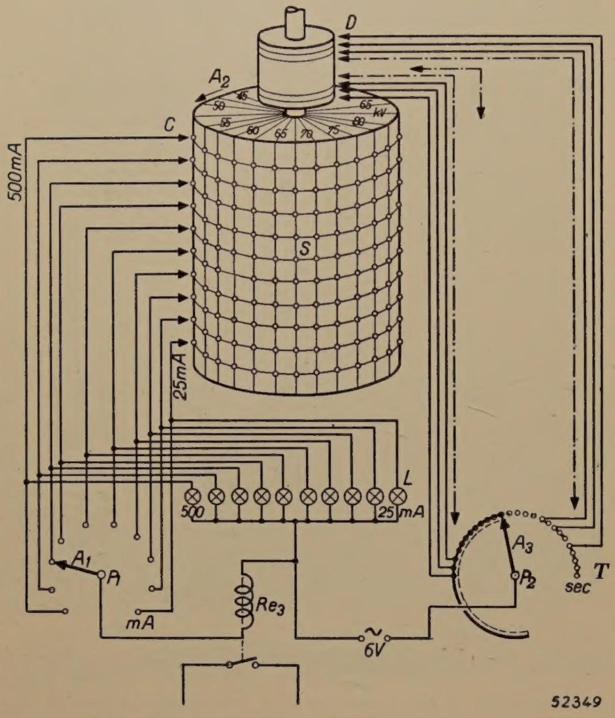

Fig. 7. Signalling system based on the auxiliary circuit shown in Fig. 5. Terminal P_1 is connected to a contact stud S by means of the ten brushes C corresponding to the ten possible mA values and selected by the arm A_1 coupled to the mA control. To each brush a lamp L is connected. When a kV value is selected by A_2 and an exposure time by A_3, all the lamps corresponding to permissible mA values will be alight, regardless of the mA value which may subsequently be chosen. The slip rings D maintain the necessary permanent connections between studs T and studs S.

Hence, the exposure can be switched on if the selected mA value is located opposite one of the illuminated mAsec figures. If, however, the selected mA value stands opposite a non-illuminated mAsec figure, the relay in the auxiliary circuit is de-energized and the X-ray tube cannot be loaded. At the same time, by the falling out of the relay a warning signal on the control table is lighted, informing the operator that the selected combination of voltage current and time is not permissible ("rating exceeded", cf. Fig. 8).

The usefulness of the above control technique is illustrated by considering the frequently occurring condition whereby the tube voltage and the number of mAsec necessary for optimal contrast and density of the radiograph are known. With the complete automatization formerly described (cf. the article mentioned in footnote [1])), only the kilovoltage and the exposure time, derived in advance from exposure tables, are to be selected. After the exposure the number of mAsec obtained is checked by means of a mAsec meter. If this number exceeds the prescribed value a shorter exposure would have been possible, with correspondingly less blurring of the image (it being presupposed that we are dealing with moving objects). With the present control technique, however, the operator, after having adjusted the tube voltage, is able to find immediately the shortest possible exposure time by turning the time control knob: he simply must

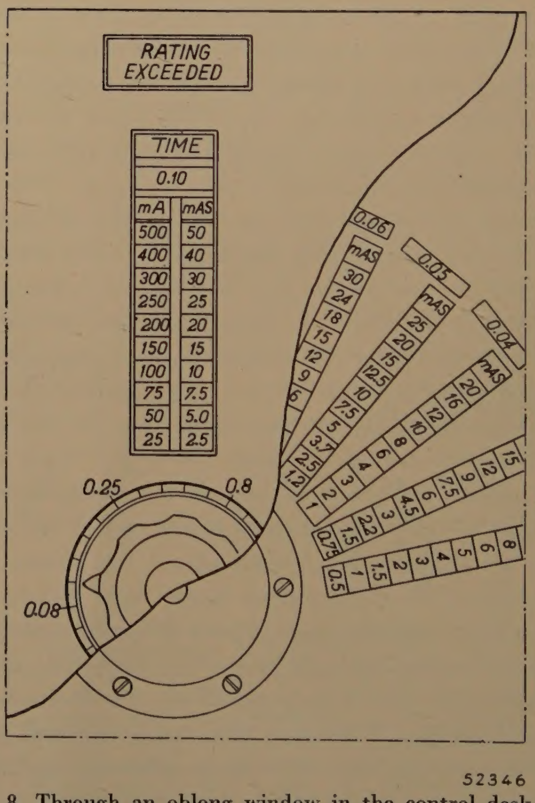

Fig. 8. Through an oblong window in the control desk the selected exposure time and mA value are read (the mA value in question being lighted in the mA column). Moreover, the operator sees at a glance which mAsec values the X-ray tube can withstand at the selected kilovoltage and exposure time: these are the numbers lighted in the mAsec column in the window. The mAsec indication is obtained by coupling to the time control handle a transparent disc carrying 26 different columns of mAsec numbers. When setting the time the correct column appears automatically in the window.

arrive at a setting where the prescribed mAsec value (or the one closest to it) will appear as the highest illuminated mAsec figure in the window. Then he has to adjust his mA control so that the illuminated mA figure appears opposite the mAsec figure in question.

Reproducibility of tube voltage and tube current

When considering the role the rating chart of the tube plays in the control technique described (Fig. 5), it will be evident that the voltage, current and exposure time obtained with the respective settings of the control handles must be accurately reproducible. This implies that their values must be largely independent of the mains supply voltage, and the high tension obtained must not be influenced by the current setting and vice versa. We shall now explain how these requirements have been met in the present apparatus.

As to the exposure time, we can chiefly refer to the description of the diagnostic apparatus repeatedly mentioned above [1] (where similar requirements concerning the reproducibility were to be fulfilled). The contactor switching the high tension transformer abruptly on and off must handle loads up to about 40 kVA. Nervetheless, no heavy moving parts and contacts must be used, since the duration of the going in and falling out must be very short, lest its consistency be impaired by friction effects. A comparatively light design of the relay is made possible by fixing the phases of making and breaking the heavy current at the moments when the pulsating high tension passes through zero value. This "isochronic" switching is obtained with a circuit in which the contactor coil is energized by means of gas-filled, grid-controlled rectifying valves. The time switch developed for our apparatus and based on the above principles may be considered as a variant of the one formerly described (cf. the article mentioned under [1])).

The tube voltage

Three chief measures are taken to make the tube voltage correctly reproducible. The first one, which is well-known, consists in compensating mains fluctuations by applying the mains voltage to various tappings on the auto transformer (T_2 in *Fig. 9*) feeding the high tension transformer. The operator, who selects the tapping by a control handle, can check the adjustment on a voltmeter which must be made to read to the nominal input voltage of 230 volts.

Another cause of variation of the adjusted tube tension is the variable voltage drop, occurring in the transformer coils, in the high tension rectifying valves, etc. due to the load current of the X-ray tube. This voltage drop varies roughly proportionally to the current. Therefore, the second measure consists in "compensating" the voltage drop variations by means of a bank of resistances in the control table. For every mA setting a resistance is provided. The resistance corresponding to the selected mA value is connected in series with the primary of the high tension transformer ($R_{3,4}$, in fig. 9). (This, of course, is equivalent to inserting a certain resistance on the high tension side.) The resistances are so adjusted in the factory that the voltage drop produced in them always supplements the voltage drop in the tube circuit to give a constant total value, regardless of the mA setting.

At the maximum possible rating of the X-ray tube (500 mA at 85 $\mathrm{kV_{peak}}$), during a short time a current of about 105 A flows through the primary circuit of the high tension transformer. As the switch for selecting the supplementary resistance has also to cope with these large currents, the transition resistance of its contacts must be extremely small (less than 0.001 ohm), lest a new and non-reproducible voltage drop be introduced. By means of a special design, making use of parallel silver contacts of the rubbing action (self cleaning) type, this has been attained with relatively small dimensions of the switch.

Apart from the voltage drop in the tube circuit another voltage drop occurs in the mains. With a current of about 100 amperes this voltage drop is far from negligible. If, for instance, the mains resistance amounts to 0.2 ohm, the voltage drop with that current is more than 20 volts, which may correspond to about 8 kV in the tube voltage. It is evident that the adjustment of the kV compensation, which is done in the factory before delivery, would be meaningless if afterwards the apparatus were to be used without taking into account the different local mains resistances that occur in different localities. Therefore, as a third correction, on installing the apparatus a padding resistance is added to the local mains resistance, giving a total of 0.4 ohm. The voltage drop caused by this constant total mains resistance is taken into account while adjusting the kV compensation in the factory and thus its effect on the adjusted tube tension is eliminated.

If the mains themselves have already a resistance larger than 0.4 ohm — which is undesirable and should be remedied if possible in order to obtain best results from the apparatus — the above described measure of "compensating" the mains resistance is impossible. Yet even in this case errors can be excluded, at least for smaller tube currents, if only the actual mains resistance is known while adjusting the apparatus in the factory: to this end the excess mains resistance (over 0.4 ohm) is subtracted from each of the resistances which are adjusted for the kV compensation. However, for the largest tube currents, for which these supplementing resistances must be very small or even zero, subtracting the surplus is not possible and one must put up with substantial deviations in the kV settings.

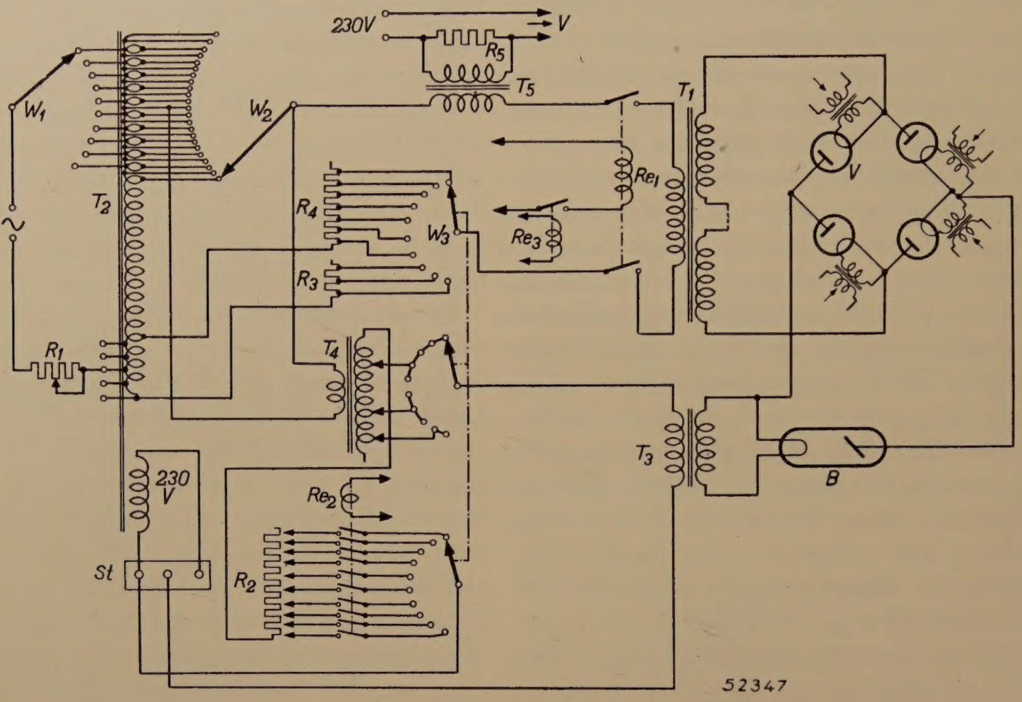

Fig. 9. Simplified circuit diagram of the DX 4 diagnostic X-ray apparatus. The high tension transformer T_1, the four rectifying valves V and X-ray tube B are connected in a conventional 4-valve rectifying system. T_1 is fed from an autotransformer T_2, connected to the A.C. mains through a mains adjusting resistance R_1. By means of the lower tappings of T_2 the apparatus is adjusted to the nominal mains voltage; the upper tappings, selected by the switch W_1, make it possible to keep the output voltage of the autotransformer constant, irrespective of mains fluctuations. W_2 is the kilovoltage selector. W_3, containing three selector arms on three decks, is the milliampere selector, which regulates in the first place the supply of transformer T_3 energizing the filament of the X-ray tube. The source of the filament current is a 230 V auxiliary winding on T_2, whose output voltage is stabilised by stabiliser St and then connected to T_3 via the filament-controlling resistance R_2 and compensating transformer T_4 (see text). The filament selector relay Re_2 is energized by the focus selector switch (not shown) connecting the correct filament resistance R_2 in the circuit. The top deck of W_3 selects the correct value of resistances R_3 and R_4, compensating for the variations of the overall voltage drop of the apparatus and the mains due to the load current. Re_1 is the high tension contactor switching the high tension circuit. Its coil is energized via a contact on the relay Re_3 whose coil is connected to the overload protection system (see text). Boosting transformer T_5 and resistance R_5 compensate for the drop in the valve filament supply due to load.

In the compensating method described it is presupposed that for a given kV setting of the transformer the voltage drop is entirely determined by the tube current. In reality it depends also on the selected voltage, due to the leakage impedance of the high tension transformer, but tests have shown that the errors caused by neglecting this effect are less than 2%.

A totally different phenomenon apt to give rise to voltage errors occurs within the rectifying valves. During heavy loads on the X-ray tube the valve filament supply voltage drops because of the voltage drop in the mains; this would cause the voltage-current characteristics of the valve to change in such a way that obtaining the desired current entails a much larger voltage loss in the valve. It is not feasible to cope with this extra voltage drop in the voltage compensation described (i.a. because of its sensitivity to minor differences in valve properties). Moreover, a large voltage loss in the valves is detrimental to their life, the valves not being designed to dissipate the resulting high wattages on the anode. In view of these consequences, a small boosting transformer (T_5 in fig. 9) is incorporated, which delivers an auxiliary voltage proportional to the current and in series with the filament voltage of the valves. The filament voltage is thus automatically maintained at the required level, in spite of variable tube loads.

The tube current

Controlling the tube current, as we have seen, is carried out by means of a variable resistance connected in series with the primary coil of the filament transformer feeding the filament of the selected focus. As the filament emission (and therefore the tube current) reacts very strongly to small variations of the filament voltage, the setting of the tube current can only be reproducible if the mains supply voltage is kept constant. Therefore the tube filament, in accordance with conventional practice, is fed from a voltage stabiliser to eliminate mains fluctuations [5]).

Other measures would not be necessary if the X-ray tube were always functioning at complete saturation of the filament emission. In practice this is not so, because in every period of the tube voltage (pulsating D.C.) low voltage values will also occur. During these parts of the period the emission drawn from the cathode has not the saturation value, but the current is limited by space charge. The higher the peak of the tube voltage, the faster the low voltage values are run through and the larger will be the portion of the period where the current attains its saturation value. Hence, the mean value of the current set by the mA control knob, when not taking any precautions, must depend to some extent on the selected tube voltage.

To eliminate this effect, i.e. to make the calibration of the mA control handle independent of the kV setting, a compensating transformer is used (T_4 in fig. 9), the output of which is added as a correction to the output of the stabiliser mentioned above. The necessary correction voltage will depend on the tube voltage and also on the current setting. Accordingly the primary voltage of the compensation transformer is varied together with the kV control, while the output is taken from various tappings on its secondary coil, the appropriate tapping being chosen by a selector arm coupled to the mA control handle.

In contrast with the compensation of the voltage drop described before, the necessary correction for the current variations is different for each X-ray tube, or rather for each focus. This follows from the fact that the necessary correction depends on space charge conditions between filament and anode. Therefore, a different series of tappings on the compensation transformer must be used for each focus. The changing of the tappings when changing the focus is performed automatically by the focus selector switch in the control table. This switch simultaneously changes the variable series resistance (R_2 in fig. 9) provided for the mA control and which, of course, must be different for each focus too. On the photograph, Fig. 4, the necessary elements are visible, assembled in four units according to the four different foci. The fifth unit (the uppermost one) is common to all filaments and is used on screening.

The measures described in the last chapter in fact have made all settings on the control table quite reproducible. Thus, on the one hand complete reliability of the overload protection of the X-ray tube is achieved, while on the other hand the quality of the radiographs obtained will be for every setting in accordance with expectations.

---

[5]) In this case a simple boosting transformer as applied for the filament voltage of the valves (see above) is not sufficient, because the filament current of the X-ray tube must be kept constant much more rigidly.

ELECTROMAGNETIC WAVES IN WAVE GUIDES

by W. OPECHOWSKI.

538.566.5:621.392.26

PART II. COAXIAL CABLES AND CIRCULAR WAVE GUIDES.

Part I of this article dealt with the theoretical principles of the propagation of electromagnetic waves in wave guides, based upon the fundamental equations of the Maxwell theory. As an example of the application of the general theory, the case of rectangular wave guides was discussed at length; in order to find electromagnetic waves capable of propagation through such wave guides, a method was followed which is based upon the superposition of plane waves.

The subject of this Part II is the propagation of electromagnetic waves in cases where the conductors have the symmetry of a cylinder with circular cross-section (coaxial cable, circular wave guide, round wire). Examples of possible modes of propagation of the waves are deduced by a direct solution of Maxwell's differential equations.

In Part I of this article ¹) we dealt with the propagation of electromagnetic waves in rectangular wave guides. As an introduction we summarized the general mathematical formulation of the principles of the Maxwell theory, followed by a fairly exhaustive discussion of the harmonic, linearly polarized plane wave as an example of a simple solution of the fundamental equations of the Maxwell theory. Finally two methods were mentioned by which these equations can be solved for a given configuration of conductors. The manner in which we found the possible modes of propagation in rectangular wave guides was an example of one of these methods, namely of that where use is made of the fact that a superposition of a number of plane waves is always a solution of the Maxwell equations (principle of superposition).

In this second part we shall deal with electromagnetic waves in coaxial cables and circular wave guides. Both these configurations of conductors have the symmetry of a cylinder with a circular cross section: a circular wave guide is formed from a coaxial cable by removing the inner conductor. The propagation of electromagnetic waves along a round wire will come in for consideration as a matter of course: such a wire is in fact what remains of a coaxial cable after the outer cylinder has been removed.

The superposition principle being absolutely general, the waves which are theoretically possible with such configurations of the conductors could also be found by suitably superposing plane waves. This, however, would mostly be rather troublesome in this case, because one would have to carry out each time the superposition of an infinite number of plane waves.

We shall therefore use these configurations of conductors to demonstrate the application of the other general method for finding the possible modes of propagation of electromagnetic waves when conductors are present.

The first step in this method is, as we said in Part I, the deduction of Maxwell's differential equations from the fundamental equations of the theory, which we have written down in the integral form.

For easy reference we write down these fundamental equations again, but without the explanation of the notation used, for which we refer to Part I pp. 15-17:

$$\oint E_l\,\mathrm{d}l = -\mu\frac{\mathrm{d}}{\mathrm{d}t}\iint H_n\,\mathrm{d}S\,, \quad . \; . \; . \; . \quad (1)$$

$$\oint H_l\,\mathrm{d}l = \sigma\iint E_n\,\mathrm{d}S + \varepsilon\frac{\mathrm{d}}{\mathrm{d}t}\iint E_n\mathrm{d}S, \quad (2)$$

$$\oint\!\!\oint H_n\,\mathrm{d}S = 0, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (3)$$

$$\varepsilon\oint\!\!\oint E_n\,\mathrm{d}S = \iiint \varrho\,\mathrm{d}\tau. \; . \; . \; . \; . \; . \; . \; . \; . \quad (4)$$

To derive the Maxwell equations in differential form from the above equations one has to begin with the choice of the system of coordinates in which one desires to obtain these equations, for with the equations (1)-(4) no particular choice of the system of coordinates is assumed ²). When a

---

¹) Philips Techn. Rev. **10**, 1948 (No. 1).

²) Maxwell's differential equations can also be given a form in which no particular choice of the system of coordinates is assumed, this being done with the aid of symbols of the vector analysis ("curl", "div", etc.).

given configuration of the conductors shows a certain symmetry then one preferably chooses a system of coordinates adapted to that symmetry. The second step is to choose suitable integration contours and surfaces in (1)-(4).

The Maxwell equations in cylindrical coordinates

It is immediately obvious that for our problem cylindrical coordinates form the system indicated. With this system the coordinate surfaces are the planes g = const. and z = const. and the cylindrical surfaces r = const.; the three kinds of coordinate surfaces are perpendicular to each other. The three components A_z, A_φ, A_r of a vector **A** in a point of the space are defined as the orthogonal projections of this vector on the normals of the three coordinate surfaces intersecting each other in this point.

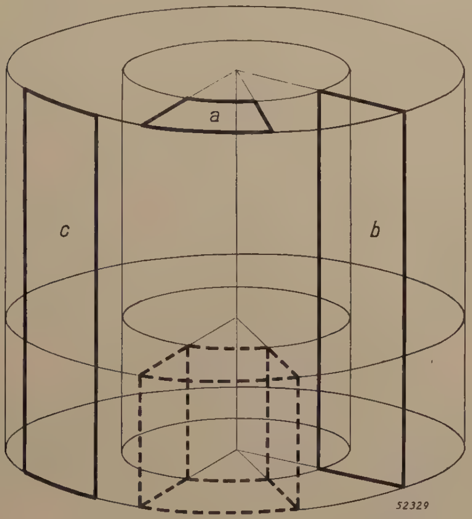

Fig. 1. Integration contours and surfaces for deduction of eqs (5)-(8) from eqs (1)-(4).

In order to derive the differential equations we choose the integration contours indicated in eq. (1)-(2) successively in the surfaces z = const., φ = const. and r = const., for instance in the manner shown in *fig. 1*. By applying eq. (1) successively to these three contours and the surfaces bounded by them and then differentiating each of the equations thus obtained with respect to the two remaining coordinates (in the case of the contour *a* in fig. 1, where the integration plane is z = const. with respect to φ and r; in the case of the contour *b* with respect to z and r; in the case of the contour *c* with respect to φ and z) one finds:

$$\frac{1}{r}\frac{\partial}{\partial r}(rE_\varphi) - \frac{1}{r}\frac{\partial E_r}{\partial \varphi} = -\mu\frac{\partial H_z}{\partial t}, \quad . \; . \quad (5a)$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -\mu\frac{\partial H_\varphi}{\partial t}, \quad . \; . \; . \; . \; . \quad (5b)$$

$$\frac{1}{r}\frac{\partial E_z}{\partial \varphi} - \frac{\partial E_\varphi}{\partial z} = -\mu\frac{\partial H_r}{\partial t} . \; . \; . \; . \; . \; . \quad (5c)$$

By performing the same operations with eq. (2) one finds further:

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\varphi) - \frac{1}{r}\frac{\partial H_r}{\partial \varphi} = \sigma E_z + \varepsilon\frac{\partial E_z}{\partial t}, \quad (6a)$$

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = \sigma E_\varphi + \varepsilon\frac{\partial E_\varphi}{\partial t}, \quad . \; . \; . \quad (6b)$$

$$\frac{1}{r}\frac{\partial H_z}{\partial \varphi} - \frac{\partial H_\varphi}{\partial z} = \sigma E_r + \varepsilon\frac{\partial E_r}{\partial t} . \; . \; . \; . \quad (6c)$$

The differential form of eqs (3) and (4) can be derived in a similar manner when the closed surface indicated in these equations is chosen as represented by dotted lines in fig. 1. One finds:

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_r) + \frac{1}{r}\frac{\partial H_\varphi}{\partial \varphi} + \frac{\partial H_z}{\partial z} = 0, \quad . \; . \; . \quad (7)$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rE_r) + \frac{1}{r}\frac{\partial E_\varphi}{\partial \varphi} + \frac{\partial E_z}{\partial z} = \frac{\varrho}{\varepsilon} . \quad . \; . \quad (8)$$

The equations (5)-(8) look rather complicated. We shall therefore at once specialize these equations to a few simple cases of practical importance.

The transverse cylindrical wave

We shall first consider the case of a purely transverse wave (radian frequency ω) being propagated in the direction of the z-axis, that is to say a wave for which E_z and H_z are zero. For the time being we shall not make any assumptions as regards the position of the conductors, expect of course that the z-axis is the axis of cylindrical symmetry. The simplest possibility now is a wave where $H_\varphi \neq 0$, $E_r \neq 0$ and all other components of the field are zero; **H** and **E** are thus at right angles to each other [3]). From (7) and (5a) it then immediately follows that H_φ and E_r do not depend upon φ; the electromagnetic field is therefore rotationally symmetrical. Further from (6a) it follows that the product rH_φ does not depend upon r. According to (8) the same holds for the product rE_r provided there are no charges in the space considered ($\varrho = 0$), which in the following we shall assume to be the case.

---

[3]) It can be proved that the equally simple transverse wave with $E_\varphi \neq 0$, $H_r \neq 0$, all other components being zero, can in no case exist, which means to say that there the field amplitudes must be zero.

The wave must therefore have the form:

$$\left.\begin{array}{l} E_r = E_r^\circ(r) \cdot e^{j(\omega t - kz + \eta)}, \\ H_\varphi = H_\varphi^\circ(r) \cdot e^{j(\omega t - kz)}, \end{array}\right\} \tag{9}$$

$$r \cdot H_\varphi^\circ(r) = C_1, \quad r \cdot E_r^\circ(r) = C_2, \tag{10}$$

where k, η, C_{10} and C_2 are constants still to be determined; $E_r^\circ(r)$ and $H_\varphi^\circ(r)$ are assumed to be real.

Now, by substituting (9) in (5b) and (6c) (these are all that remain of the eqs (5)-(8)), that is to say by writing:

$$\frac{\partial E_r}{\partial z} = -\mu \frac{\partial H_\varphi}{\partial t}, \tag{11}$$

$$-\frac{\partial H_\varphi}{\partial z} = \sigma E_r + \varepsilon \frac{\partial E_r}{\partial t}, \tag{12}$$

one easily finds

$$k^2 = \mu\omega\,(\varepsilon\omega - j\sigma), \tag{13}$$

$$k = |k|\, e^{-j\eta}, \tag{14}$$

$$\frac{C_2}{C_1} = \frac{E_r^\circ(r)}{H_\varphi^\circ(r)} = \frac{\mu\omega}{|k|}; \tag{15}$$

In writing (15) we have made use also of (10).

Eqs (13)-(14) are identical with eqs (10)-(11) in Part I, which hold for the constants k and η of a plane (damped) wave. (In this manner the use of the same symbols k and η in both cases becomes justified.) Also the expression (15) for the relation of the field amplitudes is identical with the corresponding expression for the plane wave (cf. (12) in Part I).

If one wishes to use the fact that the equations (14)-(15) apply for the plane wave (see eqs (7) and (9) in Part I)

$$\left.\begin{array}{l} E_x = E^\circ\, e^{j(\omega t - kz + \eta)}, \\ H_y = H^\circ\, e^{j(\omega t - kz)}, \end{array}\right\} \tag{16}$$

then the validity of the same equations for the wave (9) can be proved without any calculation, for the eqs (11)-(12) are identical with the eqs (5b) and (6a) in Part I, which are satisfied by the plane wave (16).

The solution of the Maxwell equations (5)-(6) characterized by the eqs (9)-(10) and (13)-(15) is sometimes called the (damped) transverse cylindrical wave. In order to characterize this wave completely we must still determine the constants C_1 and C_2.

Now it is easily understood that in an unbounded homogeneous medium C_1 and C_2 are zero; in other words, a transverse cylindrical wave cannot be propagated in such a medium. The fact is that since $E_z = 0$ and ϱ and ε are finite the right-hand side of eq. (2) must be 0. This means that the line integral in the left-hand side is also zero. When one takes a circle $r =$ const. as integration contour this line integral has the value $2\pi r H_\varphi^\circ$. Thus we obtain $2\pi r H_\varphi^\circ = 0$ and from that, owing to (10), $C_1 = 0$. From (15) it then follows that also $(C_2) = 0$. Therefore the amplitudes of the transverse cylindrical wave become in this case zero.

This manner of reasoning is no longer applicable, however, when we have along the z-axis a perfectly conducting wire (radius a). E_z still remains zero it is true, but since $\sigma = \infty$ we may have the product $\sigma E_z \neq 0$, implying a surface current along the surface of the wire.

In order to determine the constants C_1 and C_2 in this case, we shall employ the boundary conditions which have to be satisfied by the electromagnetic field at the boundary surface between two media. These boundary conditions were briefly discussed in Part I and formulated in eqs (21)-(22) for the case where one of the two media is a perfect conductor.

When the boundary surface is a cylindrical surface, as it is here, then on the basis of these equations we have at the boundary surface:

$$E_\varphi = 0, \qquad H_r = 0, \tag{17}$$

$$\varepsilon E_r = s, \qquad H_\varphi = i, \tag{18}$$

where s and i denote the density of the surface charge and of the surface current respectively (the amplitudes of these quantities will be denoted by s° and i°).

From (10) and (18) it now follows that:

$$C_1 = a i^\circ, \qquad \varepsilon C_2 = a s^\circ. \tag{19}$$

The boundary condition (17) is already satisfied because for a transverse cylindrical wave E_φ and H_r are everywhere zero.

When the medium in which the wire is situated is a perfect insulator ($\sigma = 0$) then from eqs (13), (15) and (19) it is easy to convince oneself that the ratio i°/s° is equal to the phase velocity $v = \omega/k$ of the cylindrical wave.

In what follows we shall need also an equation to indicate the relation between $H\varphi$ and the total surface current $I = 2\pi a i$ along the wire. This equation follows immediately from (10) and (19):

$$H_\varphi\,(z, r, t) = \frac{I\,(z,t)}{2\pi r}. \tag{20}$$

In the region of very high frequencies a wire therefore plays only the role of a cylindrical boundary surface making it possible for a transverse cylindrical wave to be propagated in a certain direction. Actually, owing to the imperfect conductivity of the wire, the wave cannot be exactly transverse, that is to say E_z cannot be exactly zero. However, the higher the frequency, the smaller is the error resulting from putting $E_z = 0$.

In a coaxial cable it is of course possible for the transverse cylindrical wave to be propagated, since on the inside of the outer cylinder (radius b) the boundary conditions are automatically satisfied. There the total surface current is also $I(z,t)$ but the direction of the current is opposed to that of the surface current on the wire, i.e. on the inner conductor (radius a) of the cable.

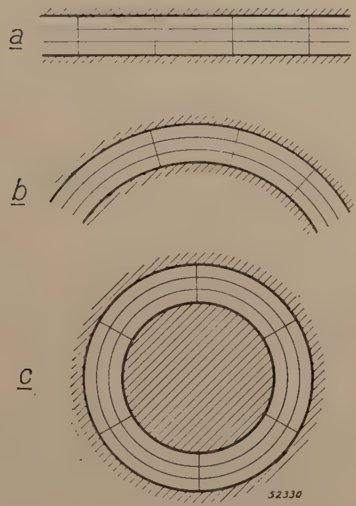

Fig. 2. The manner in which the magnetic lines of force (concentric circles) and of electric lines of force (straight lines) for a transverse cylinder wave in a coaxial cable (*c*) can be deduced from that for a transverse plane wave between two parallel flat planes (*a*) by gradually transforming the latter into curved surfaces (*b*). In all three cases the direction of propagation of the wave is perpendicular to the plane of the drawing.

The lines of force of the **E** and **H** fields are diagrammatically represented in *fig. 2c*. We take this opportunity to show also in figs 2*a* and *b* how the trend of the lines of force for the transverse cylindrical wave in a coaxial cable can be graphically "deduced" from that for the transverse wave between two flat planes (discussed in Part I) by gradually curving the planes. This is often a useful way of arriving at an idea of the trend of the lines of force also for a non-transverse wave without working out the solution of the Maxwell equations; this particularly applies in cases where an exact solution is practically impossible owing to mathematical difficulties. Of course the lines of force for a certain form of conductors must already be known.

If we now imagine the inner conductor to be removed from a coaxial cable then, in order to determine C_1 and C_2 of the transverse cylindrical wave within the space bounded by the outer cylinder, we can apply without any alteration the reasoning followed for the case of an unbounded homogeneous medium. In this manner we arrive again at the conclusion that $C_1 = C_2 = 0$. In other words, a transverse cylindrical wave cannot be propagated in a circular wave guide.. This result is a special case of the general theorem which we formulated in Part I, viz. that a purely transverse electromagnetic wave cannot be propagated in any wave guide the cross section of which is bounded by one single continuous closed curve. (The cross section of a coaxial cable is bounded by two closed curves; therefore the hypothesis does not hold in that case.)

Deduction of the "cable equations"

It can easily be proved that eqs (11)-(12) are equivalent to the well-known "cable equations" for the tension V and the current I in a coaxial cable. The proof is as follows. The current $I(z,t)$ is defined by (20) and the voltage $V(z,t)$ between the outer and inner conductors by

$$V(z,t) = \int_A^B E_l \, \mathrm{d}l \,, \quad . \; . \; . \; . \; . \; . \; . \quad (21)$$

the path of integration lying entirely in the plane $z =$ const., whilst its end points A and B are situated on the outer and inner conductors; apart from this the path of integration is arbitrary.

It is easy to understand that in this manner the tension $V(z,t)$ is unambiguously defined.

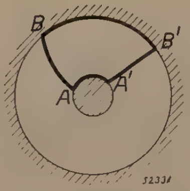

Fig. 3. Integration contour $ABB'A'A$ for proving that the tension V between the two conductors in the coaxial cable is unambiguously defined by eq. (21).

To make this clear it is noted that the integral $\int E_l \mathrm{d}l$ along the closed path $ABB'A'A$ indicated in *fig. 3* must be equal to zero, regardless where the points A, A' and B, B' lie respectively on the inner conductor and on the outer conductor and what the shape of the curves AB and $A'B'$ may be in the plane z = const., for since $H_z = 0$ the magnetic flux through the part of the plane z = const. bounded by $ABB'A'A$ is always zero. In accordance with the law of induction (eq. (1)) it follows from this that the said integral is indeed zero. Now:

$$\oint_{ABB'A'A} E_l\,\mathrm{d}l = \int_A^B E_l\,\mathrm{d}l + \int_B^{B'} E_l\,\mathrm{d}l + \int_{B'}^{A'} E_l\,\mathrm{d}l + \int_{A'}^{A} E_l\,\mathrm{d}l$$

applies.

Since $E_\varphi = 0$, the second and fourth integrals in the right-hand side of this equation are zero, i.e.:

$$0 = \int_A^B E_l\,\mathrm{d}l + \int_{B'}^{A'} E_l\,\mathrm{d}l,$$

or

$$\int_A^B E_l\,\mathrm{d}l = \int_{A'}^{B'} E_l\mathrm{d}l,$$

from which the unambiguity of the definition (21) immediately follows.

When we substitute in eqs (11)-(12) the expression (20) for H_φ and then integrate these equations with respect to r in the manner indicated in the definition (21) of V, then after a little manipulation we find:

$$-\frac{\partial V}{\partial z} = L\frac{\partial I}{\partial t}, \quad \ldots \ldots \quad (22)$$

$$-\frac{\partial I}{\partial z} = GV + C\frac{\partial V}{\partial t}, \quad \ldots \quad (23)$$

in which

$$L = \frac{\mu}{2\pi}\cdot\ln(b/a), \quad C = \frac{2\pi\varepsilon}{\ln(b/a)}, \quad G = \frac{2\pi\sigma}{\ln(b/a)}, \quad (24)$$

where L, C and G denote respectively the series self-inductance, the shunt capacity and the shunt conductance per metre of the cable.

The conductors of the cable being assumed to be perfectly conducting, the term RI (R = resistance per metre) usually present in the right-hand side of equation (22) is of course absent, for this term describes approximately the losses in the conductors due to the development of Joule heat.

The cable equations (22)-(23), deduced here for a coaxial cable from the Maxwell equations [4]), also hold strictly for any configuration of two arbitrary conductors allowing the propagation of a purely transverse wave (naturally the expressions for L, C and G vary from case to case). We shall not give here the proof of this more general statement, the plausibility of which is so obvious.

In this connection we would make the following remark. In the general case the voltage V is defined also by (21). The unambiguity of this definition is based, as we have seen, only upon two facts: 1) that the component of **H** in the direction of propagation of the wave is zero, 2) that the tangential component of **E** at the surface of the conductors is also zero. The former follows from the transverse nature of the wave, the latter from the boundary condition. The current I has in the general case the same meaning as in the case of the coaxial cable, namely that of the total surface current along one of the conductors.

Rotationally symmetrical E and H waves [5])

In the foregoing it has been shown that a purely transverse cylindrical wave can only be propagated when there is a conductor along the axis of symmetry. Neither in free space nor in a circular wave guide can such a wave be propagated. The essential condition for the possibility of propagation of an electromagnetic wave in a circular wave guide is, therefore, that at least one of the field vectors **E** and **H** must have a component in the direction of propagation, i.e. in the direction of the axis of symmetry. The two simplest kinds of waves are then:

1) Waves where $H_\varphi \neq 0$, $E_r \neq 0$, $E_z \neq 0$, all other components of the **E** and **H** fields being zero. In accordance with what we have stated when dealing with rectangular wave guides, these waves and in general the waves for which $H_z = 0$ are called the "E-waves" (or "TM-waves").

2) Waves where $E_\varphi \neq 0$, $H_r \neq 0$, $H_z \neq 0$, all other components being zero. These waves are examples of "H-waves" (or "TE-waves").

In the case of each of these waves **E** and **H** are thus at right-angles to each other. From eqs (5)-(8) it now follows that with these two kinds of waves the **E** and **H** fields are not dependent upon φ; these waves are therefore rotationally symmetrical.

In Part I it has already been stated that in wave guides waves which are not purely transverse necessarily have a cut-off frequency; if the frequency of such a wave is lower than the cut-off frequency in a given wave guide then it is impossible for the wave to be propagated in that wave guide. In the following we shall briefly outline how one arrives at the cut-off frequency of the rotationally symmetrical E- and H-waves. We shall confine the discussion of these waves mainly to this point.

---

[4]) We could, of course, have derived the cable equations (22)-(23) directly from the fundamental equations (1)-(2) without employing the differential equations (11) and (12) as intermediate stage.

[5]) The first publications on the theory of circular wave guides appeared in 1936: J. R. Carson, S. P. Mead and S. A. Schelkunoff, Bell Syst. Techn. J. **15**, 310-333, 1936; W. D. Barrow, Proc. Inst. Radio Engrs **24**, 1298-1328, 1936; L. Brillouin, Rev. Gén. d'Electr. **40**, 227-239, 1936. The practical importance of wave guides in general was first pointed out in the article by Barrow quoted above and in an article by G. C. Southworth, Bell Syst. Techn. J. **15**, 284-309, 1936, which latter article is complementary to the publication by Carson, Mead and Schelkunoff.

We assume that in both cases all components of the **E** and **H** fields have the same factor $e^{j(\omega t - hz)}$ expressing wave propagation in the z-direction, thus

$$H_\varphi = H_\varphi{}^\circ(r) \cdot e^{j(\omega t - hz)},$$

etc.

It will be found that as a rule the field amplitudes are complex numbers, which means to say that the various components of the **E** and **H** fields do not oscillate in phase [6]).

Rotationally symmetrical E-waves

By substituting in (5)-(6) H_φ, E_r, E_z, which have the form just mentioned, and taking the remaining components as zero, after an easy reduction one obtains the following equations for rotationally symmetrical E-waves:

$$\frac{\mathrm{d}^2 H_\varphi{}^\circ}{\mathrm{d}\varrho^2} + \frac{1}{\varrho}\frac{\mathrm{d}\,H_\varphi{}^\circ}{\mathrm{d}\varrho} + \left(1 - \frac{1}{\varrho^2}\right) H_\varphi{}^\circ = 0, \quad (25)$$

$$E_z{}^\circ = \frac{\sqrt{k^2 - h^2}}{\sigma + j\varepsilon\omega}\,\frac{1}{\varrho}\,\frac{\mathrm{d}}{\mathrm{d}\varrho}\,(\varrho\,H_\varphi^\circ)\,, \quad . \; . \; . \; . \quad (26)$$

$$E_r{}^\circ = \frac{jh}{\sigma + j\varepsilon\omega}\,H_\varphi{}^\circ, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (27)$$

in which

$$\varrho = r\sqrt{\mu\omega\,(\varepsilon\omega - j\sigma) - h^2}\,;$$

if we denote the first term in the expression below the radical by k^2 (see (13)) then this expression can be abbreviated to read:

$$\varrho = r\sqrt{k^2 - h^2}\,. \quad . \; . \; . \; . \; . \; . \quad (28)$$

The constant h will presently be determined by the boundary condition.

From these formulae it appears that E_r and E_z are unambiguously determined by H_φ.

Now the equation for H_φ is an equation very well known in mathematics, viz. the equation for the Bessel functions of the first order. Incidentally it may be stated that by replacing $1/\varrho^2$ by n^2/ϱ^2 ($n = 0, 1, 2 \ldots\ldots$) one obtains the equation for the Bessel functions of the nth order. Among the Bessel functions two kinds are distinguished. The Bessel function of ϱ of the nth order "of the first kind" [7]) is denoted by $J_n(\varrho)$. The Bessel functions $\mathrm{J}_0(\varrho)$ and $\mathrm{J}_1(\varrho)$, the only ones with which we shall be concerned here, are shown in *fig. 4;* these two functions are related by the following expressions:

$$\frac{\mathrm{d}\mathrm{J}_0}{\mathrm{d}\varrho} = -\mathrm{J}_1 \quad \text{and} \quad \frac{1}{\varrho}\frac{\mathrm{d}}{\mathrm{d}\varrho}\,(\varrho\,\mathrm{J}_1) = \mathrm{J}_0\,. \quad . \quad (29)$$

Thus according to (25)-(27) and (29) the amplitudes of the rotationally symmetrical E-wave are given by

$$\left.\begin{aligned} H_\varphi{}^\circ = A\mathrm{J}_1(\varrho), \qquad E_z{}^\circ &= A\,\frac{\sqrt{k^2 - h^2}}{\sigma + j\varepsilon\omega}\,\mathrm{J}_0(\varrho), \\ E_r{}^\circ &= A\,\frac{jh}{\sigma + j\varepsilon\omega}\,\mathrm{J}_1(\varrho)\,, \end{aligned}\right\} \quad (30)$$

in which A is an arbitrary constant.

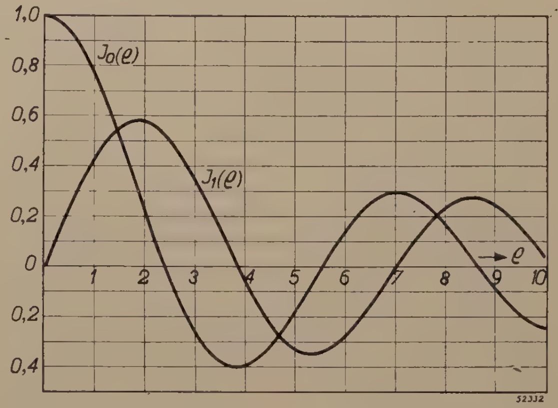

Fig. 4. Graphic representation of the Bessel functions $\mathrm{J}_0(\varrho)$ and $\mathrm{J}_1(\varrho)$.

We shall from now on assume that the wall of the wave guide is perfectly conducting. According to the boundary conditions (21) in Part I the tangential component E_t of **E** and the normal component H_n of **H** must therefore be zero at the wall.

Given that the wave guide has a radius b, the boundary condition $E_t = 0$ requires that $E_z(b) = 0$, that is to say

$$\mathrm{J}_0(b\sqrt{k^2 - h^2}) = 0. \quad . \; . \; . \; . \; . \; . \quad (31)$$

The boundary condition $H_n = 0$ is automatically satisfied, for the radial component of **H** here is everywhere zero.

From (31) one can determine h. The function $\mathrm{J}_0(\varrho)$ has an infinite number of zero points ϱ_1, $\varrho_2, \ldots\ldots$, some of which are shown in fig. 4: ϱ_1 has the value 2.405, $\varrho_2 = 5.520$, $\varrho_0 = 8.654$, and so on. For the quantity h we therefore derive from (31) an infinite number of equations of the form

$$h_p = \sqrt{k^2 - (\varrho_p/b)^2} \qquad (p = 1, 2, \ldots), \quad (32)$$

---

[6]) In the case of the transverse cylindrical wave (9) we were able to assume that the field amplitudes are real because we explicitly introduced the phase difference η, which we have not done here.

[7]) We should come across the Bessel function of the "second kind" when discussing the E and H waves in a coaxial cable.

where the radius b of the wave guide is taken as given.

For air (or vacuum) we may put $\sigma = 0$. From (13) it follows that k^2 is then real, viz. $k^2 = \varepsilon\mu\omega^2 = \omega^2/c^2$.

Eq. (32) thus becomes:

$$h_p = \sqrt{(\omega/c)^2 - (\varrho_p/b)^2} \quad (p = 1, 2, \ldots). \quad (33)$$

Now only real values of h are of importance for our problem, because when h is imaginary this means that the wave is not propagated (this has been explained in Part I). From (33) we see that h can only be real when

$$(\omega/c)^2 - (\varrho_p/b)^2 \geqq 0\,,$$

that is to say when

$$\nu > \varrho_p c/2\pi b$$

or

$$\lambda < 2\pi b/\varrho_p.$$

Thus, just as in Part I in the case of rectangular wave guides, we meet here again the concepts "cut-off frequency" ν_c and "cut-off wavelength" λ_c:

$$\lambda_c = c/\nu_c = 2\pi b/\varrho_p. \quad \ldots\ldots \quad (34)$$

Only rotationally symmetrical E-waves with a wavelength $\lambda < \lambda_c$ can be propagated in a circular wave guide. A certain λ_c corresponds to every value of ϱ_p. The greatest cut-off wavelength for rotationally symmetrical E-waves is therefore

$$\lambda_c^{(1)} = 2\pi/2.405 = 2.61\,b, \quad \ldots \quad (35)$$

and this is found to be also the greatest wavelength that a wave can have in a circular wave guide.

Denoting the wavelength of the wave in the wave guide by λ_z (cf. the analogous notation in Part I), from (33) and (34) it follows that:

$$\lambda_z^{(p)} = 2\pi/h_p = 1/\sqrt{(1/\lambda)^2 - (1/\lambda_c)^2}\,. \quad \ldots \quad (36)$$

The z-dependence of the rotationally symmetrical E-wave is therefore given by

$$e^{-jh_p z} = e^{-2\pi jz\sqrt{(1/\lambda)^2 - (1/\lambda_c)^2}}. \quad \ldots\ldots \quad (37)$$

Precisely the same formula has been deduced in Part I for a rectangular wave guide (eq. (33) Part I).

For easy reference we again write down here the full expression (in the non-complex notation) for rotationally symmetrical E-waves in a wave guide with perfectly conducting wall:

$$\left.\begin{aligned}
E_z &= (A/\varepsilon\omega)\sqrt{(\omega/c)^2 - {h_p}^2}\cdot \mathrm{J}_0(\varrho)\cdot \sin(\omega t - h_p z),\\
E_r &= (Ah_p/\varepsilon\omega)\cdot \mathrm{J}_1(\varrho)\cdot \cos(\omega t - h_p z),\\
H_\varphi &= A\cdot \mathrm{J}_1(\varrho)\cdot \cos(\omega t - h_p z),\\
E_\varphi &= H_r = H_z = 0,
\end{aligned}\right\} \quad (38)$$

in which

$$\varrho = r\sqrt{(\omega/c)^2 - {h_p}^2}\,; \quad \ldots\ldots \quad (39)$$

h_p is defined by (33).

Corresponding to a certain value of $p = 1, 2, \ldots$ there is an E-wave of a certain form, for p indicates the number of coaxial cylindrical surfaces in the wave guide for which $|E_z|$ is the maximum and E_r and H_φ are zero, when the axis of symmetry is also considered as a (degenerated) cylindrical surface; this statement can be verified with the aid of fig. 4. The electrical lines of force for a wave corresponding to ϱ_1 are shown in *fig. 5*, in the part lying between the two horizontal dotted lines representing the wall of the wave guide. The magnetic lines of force lie in planes z = const. and are concentric circles. The density of the surface current and surface charge on the wall of the wave guide follows from the boundary conditions (18).

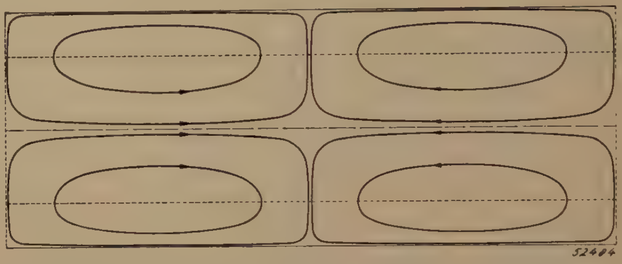

Fig. 5. Representation of the electrical lines of force for a rotationally symmetrical E-wave (corresponding to the smallest root of the eq. $\mathrm{J}_0(\varrho) = 0$) and of the magnetic lines of force for a rotationally symmetrical H-wave (corresponding to the smallest root of the eq. $\mathrm{J}_1(\varrho) = 0$ differing from zero) in a circular wave guide. In both cases the lines of force are determined by the equation

$$\varrho\cdot \mathrm{J}_1(\varrho)\cdot \sin 2\pi\,\zeta = \text{constant.}$$

The variable ζ is given by $\zeta = z/\lambda_z$, in which λ_z is the wavelength of the wave in the wave guide (see eq. (36)) and the z-axis coincides with the axis of symmetry (dot-dash line) of the wave guide; the variable ϱ is proportional to the distance r from the axis of symmetry (see eq. 28)).

The plane of the drawing is thus an r-z-plane (or φ = const.). In the case of H-waves the fully drawn horizontal straight lines represent the intersections of the wall of the wave guide with the plane of the drawing. The contours indicate the magnetic lines of force. The same contours also represent the electrical lines of force in the case of the E-wave, provided the wall of the wave guide is imagined as being removed to the dotted horizontal straight lines.

The solution (30) of eqs (5)-(6) forms also the basis of the theory of the skin effect in a circular wire, assuming σ to be finite inside the wire and zero outside it. Thus the variable ϱ is complex. On the surface of the wire one then has to try to satisfy the boundary conditions, which in this case lead to rather complicated formulae, since neither of the two media is a perfect conductor.

Rotationally symmetrical H-waves

Rotationally symmetrical H-waves can be dealt with in exactly the same manner. A comparison of (5) with (6) shows at once that the expressions for the amplitudes of a rotationally symmetrical H-wave can be obtained by replacing $H_\varphi{}^\circ$, $E_z{}^\circ$, $E_r{}^\circ$, $(\sigma + j\varepsilon\omega)$ and $(-j\mu\omega)$ in (30) respectively by $E_\varphi{}^\circ$, $H_z{}^\circ$, $H_r{}^\circ$, $(-j\mu\omega)$ and $(\sigma + j\varepsilon\omega)$. The constant k^2 is not changed by this substitution since according to (13)

$$k^2 = (-j\mu\omega)\,(\sigma + j\varepsilon\omega).$$

This time, therefore, the boundary condition $E_t = 0$ does not lead to eq. (31) but to the following equation for the determination of h:

$$\mathrm{J}_1(b\sqrt{k^2 - h^2}) = 0 \quad . \; . \; . \; . \quad (40)$$

Since $E_\varphi{}^\circ$ and $H_r{}^\circ$ (just as $H_\varphi{}^\circ$ and $E_r{}^\circ$ in the case of the E-waves) depend in like manner upon r, the boundary condition $H_n = 0$ is automatically satisfied. Eqs (34)-(37) hold, of course, also here, the only difference being that ϱ_p is no longer determined by the eq. $J_0\,(\varrho) = 0$ but by the eq. $J_1\,(\varrho) = 0$. The full expression for the rotationally symmetrical H-waves in the case when $\sigma = 0$ is obtained by replacing E_z, E_r, H_φ and ε in (38) respectively by H_z, H_r, E_φ and $(-\mu)$. The magnetic lines of force for a simple rotationally symmetrical H-wave are also given in fig. 5.

There is a peculiarity about the rotationally symmetrical H-waves to which we would draw attention. On the wall of the wave guide the magnetic field follows the direction of the z-axis ($H_r(b)$ being zero) and hence the surface current flows at right-angles to the z-direction: the lines of flux are circles perpendicular to the z-axis [8]). This is a striking example of the fact that at high frequencies there is no direct relation between the direction of the current and the direction of propagation of the electromagnetic wave along the conductor, thus also of the electromagnetic energy.

Apart from the rotationally symmetrical E- and H-waves considered above, in circular wave guides E- and H-waves may be propagated of which the **E** and **H** vectors depend also upon the angle coordinate φ. The form of these waves is deduced in a similar manner from equations (5)-(8), though the calculations in this case are somewhat more complicated; we shall not go into that further here.

Finally we would observe that in many technically important problems relating to the propagation of microwaves a rigorous solution of the Maxwell equations involves great mathematical difficulties; in the foregoing we have considered only exceptionally simple configurations of conductors. Nevertheless the methods that we have outlined are useful when seeking approximative solutions of more complicated problems. This will become evident when the technique of wave guides is dealt with in further articles to be published in this journal.

APPENDIX:

About the relation between wave guides and cavity resonators

If a circular wave guide is closed by applying a metal wall in any two planes perpendicular to the axis of symmetry then a (circular) cavity resonator is obtained. The characteristic modes of oscillation of such cavity resonators have already been dealt with in this journal [9]) [10]). In the method of treatment applied there, however, deductions were limited to that of the distribution of the current density in the walls of a circular, flat cavity resonator. We shall here outline briefly how one arrives at the expressions for the electromagnetic field in the space bounded by the walls of a circular cavity resonator (not necessarily flat).

Obviously in a stationary condition the electromagnetic field must bear the character of a standing wave. In the case of a standing electromagnetic wave the electrical and magnetic lines of force are not dependent upon the time; only the strength of the field changes with time (in the most simple case sinusoidally). In other words the expression for **E** or **H** for a standing wave must be a product of a function depending only upon time and a function depending only upon the place.

A standing wave (in a circular cavity resonator) can be imagined as originating from a superposition of two running waves in a circular wave guide having opposite directions of propagation but otherwise absolutely identical. We shall explain this further with the example of the rotationally symmetrical E-waves. The superposition referred to means that to the expressions (38) for the field components one has to add the expressions obtained from (38) when replacing h_p by $(-h_p)$. In this way we obtain:

$$\left.\begin{aligned}
E_z &= (2A/\varepsilon\omega)\sqrt{(\omega/c)^2 - h_p{}^2}\,.\,\mathrm{J}_0(\varrho)\,.\cos h_p z\,.\sin \omega t,\\
E_r &= -(2Ah_p/\varepsilon\omega)\,.\,\mathrm{J}_1(\varrho)\,.\sin h_p z\,.\sin \omega t,\\
H_\varphi &= 2A\,.\,\mathrm{J}_1(\varrho)\,.\cos h_p z\,.\cos \omega t,\\
E_\varphi &= H_z = H_r = 0.
\end{aligned}\right\}\quad (41)$$

---

[8]) Related to this property of the rotationally symmetrical H-waves is the fact that an unlimited increase of the frequency ν is accompanied by a decrease of the energy losses in the wall of the wave guide right down to zero; for all other H and E waves the energy losses for $\nu = \infty$ become infinitely large.

[9]) C. G. A. von Lindern and G. de Vries, Flat cavities as electrical resonators, Philips Techn. Rev. **8**, 149-160, 1946 (No. 5).

[10]) G. de Vries, Electromagnetic cavity resonators, Philips Techn. Rev. **9**, 73-84, 1947 (No. 3).

This is indeed a standing wave, for **E** and **H** are products of $\sin \omega t$ and $\cos \omega t$, respectively, and a function depending only upon the place.

The expressions (41) will represent a possible mode of oscillation of the **E** and **H** fields in a circular cavity resonator when the boundary conditions are satisfied at the two new walls (transforming a circular wave guide into a circular cavity resonator). If the two walls are introduced at $z = 0$ and at $z = l$ then at $z = 0$ the boundary condition $E_t = 0$ is automatically satisfied, since $E_r = 0$ for $z = 0$. To obtain $E_r = 0$ also at $z = l$, we must have $h_p l = 0$, and from this it follows that:

$$h_p l = \pm \pi q \; (q = 0,\ 1,\ 2\ldots). \quad . \quad . \quad . \quad . \quad . \qquad (42)$$

The boundary condition $H_n = 0$ is automatically satisfied because H_z is everywhere equal to zero.

Now the condition (42) agrees with eq. (33) only when $\omega = \omega_{pq}$, where

$$\left(\frac{\omega_{pq}}{c}\right)^2 = \left(\frac{\pi q}{l}\right)^2 + \left(\frac{\varrho_p}{b}\right)^2. \qquad \begin{pmatrix} p = 1, 2, 3\ldots \\ q = 0, 1, 2\ldots \end{pmatrix} \qquad (43)$$

From eq. (43) it follows that a circular cavity resonator cannot oscillate at every frequency. Among the rotationally symmetrical standing waves (41) only those corresponding to the characteristic frequencies $\nu_{pq} = \omega_{pq}/2\pi$ represent a possible mode of oscillation.

According to (43) and (34) the lowest characteristic frequency ($q = 0$, $p = 1$) is equal to the lowest cut-off frequency of running E-waves in a circular wave guide.

$$\nu_{10} = \varrho_1 c/2\pi b. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (44)$$

Not only the "standing E-wave" corresponding to (44) — here we introduce a term which is in line with the terminology used in the case of wave guides — but also all other standing E-waves with $q = 0$ have the property that their characteristic frequencies are independent of the length l of the cavity resonator; this follows from (43). From (41) and (42) it follows moreover that with these standing waves the electrical lines of force always run parallel to the axis of symmetry. In the case of "standing H-waves", which we shall not deal with here, the characteristic frequency is in no case independent of the length of the cavity resonator.

The electrical lines of force in the case of a standing E-wave ($q = 1, p = 1$) in a circular cavity resonator can be seen from fig. 5, provided the two new walls closing the wave guide on the left and the right are applied at suitably chosen places. By "suitably" we mean here a choice which satisfies the boundary conditions. A vertical straight line, which in fig. 5 would represent each of the new walls, must therefore exactly halve the contours. (In a similar manner the magnetic lines of force can be read from fig. 5 for the case of a standing H-wave in a circular cavity resonator.) Whereas, however, in the case of the wave guide fig. 5 was "an instantaneous recording", here this figure represents lines of force not changing with the time.

The current density i in the walls of the cavity resonator can also be easily calculated from eqs (41)-(43) if reference is made to the boundary condition $H_t = i$. For the "bottom" ($z = 0$) or the "cover" ($z = l$) of the cavity resonator we obtain in this manner, for $q = 0$,

$$i = H_\varphi = 2A \,.\, J_1\,(r\varrho_p/b) \,.\, \cos \omega t \,. \quad . \quad . \quad . \qquad (45)$$

Fig. 2 of the article quoted in footnote [10]) is a representation of the current density distribution as obtained from (45) for $p = 1$ and $p = 2$.

Of course the mathematical expressions for electromagnetic waves in cavity resonators can also be deduced directly from the Maxwell equations, thus without employing the expressions for running waves.

THE ORIGIN OF THE GIORGI SYSTEM OF ELECTRICAL UNITS

by W. de GROOT.

537.71

This article gives an outline of the origin of the system of fundamental formulae in electricity and magnetism, the subject being dealt with in such a way as to give an insight into the Giorgi system of electrical units. These units are explained and particular attention is paid to what is called rationalization. This is followed by a review dealing historically with the c.g.s. systems, the practical system, the absolute and the international practical units. Finally the realization of the absolute ampere, ohm and volt is briefly described.

Introduction

The following is an outline of the development of the fundamental formulae in electromagnetism, given in such a way as to throw some light upon the Giorgi or M-K-S system of electrical units.

The unit of charge

Systems of formulae used in electro-magnetism are usually based upon the Coulomb law:

$$K = \alpha \frac{Q_1 Q_2}{r^2},$$

where K is the mutual force between the charges Q_1 and Q_2 at a distance r between each other, and α is a constant that has still to be determined [1]. Once it has been decided in what units K and r are to be measured, then by ascribing a certain value to α one arrives at a definition for a unit of charges.

One may also regard the above formula as a particular case of a general law according to which

$$\mathbf{K} = Q_2 \cdot \mathbf{E},$$

where **E** represents the field strength vector [2]. In an electrostatic field, for a charge Q_1 carried along a path s from point *1* to point *2* we have the formula:

$$\frac{1}{Q_1}\int_1^2 K_s \,\mathrm{d}s = \int_1^2 E_s \,\mathrm{d}s = V_1 - V_2,$$

where K_s and E_s represent the components of the force and of the field strength respectively in the direction of s. This holds regardless of the path followed by the charge from point *1* to point *2*. V_1 and V_2 are the electrostatic potentials in the points *1* and *2* (already determined except for a constant). This determines the unit of potential difference, since the unit in which the energy $\int K_s \,\mathrm{d}s$ will be measured is already fixed.

An important step in the theory is the law of the superposition of electric fields: **E** is the vector sum of the separate **E**'s originating in each of the charges. As a consequence, when we integrate in vacuum the component E_n of **E** across a closed surface A according to the normal of that surface and thus construct $\oiint E_n \mathrm{d}A$, the result is proportional to the total charge Q enclosed by that surface:

$$Q = \iiint \varrho \,\mathrm{d}\tau \backsim \oiint E_n \,\mathrm{d}A,$$

where ϱ represents the spatial charge density (charge per unit volume) whilst $\mathrm{d}\tau$ is a volume element within the enclosed surface. The value of the constant of proportionality (which appears to be $4\pi\alpha$) is directly related to the choice of the unit of charge.

Following the example of Maxwell, the same fact can also be expressed by introducing a new vector **D** proportional to **E** and then requiring that the corresponding surface integral $\oiint D_n \mathrm{d}A$ shall be not only proportional but also equal to Q:

$$\oiint D_n \,\mathrm{d}A = Q. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

The integral of D_n across a part of a surface bounded by a closed curve is called the electric flux Ψ through that surface

$$\Psi = \iint D_n \,\mathrm{d}A.$$

Equation (1) then expresses the fact that the total electric flux Ψ through a closed surface is equal to the enclosed charge Q.

The relation between **D** and **E** can be represented by writing:

---

[1] The Coulomb law has not been very accurately verified by direct observation of the force K. Later, it has been confirmed very accurately by indirect experiments, for instance by Cavendish, where it was proved that a charged body placed inside a closed conductive envelope lost all its charge to the envelope when contacted with it. See, for instance, J. H. Jeans, The mathematical theory of electricity and magnetism, Cambridge 1923, 4th edition, pp. 37-38.

[2] From now onwards vectors, such as **E**, will be printed in heavy type and the magnitude of a vector, such as E, in italics.

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where the magnitude of ε again depends upon the units chosen.

The advantage of proceeding in this way is that account is also taken of the phenomena occurring when the dielectric is not an empty space but is wholly or partly filled with an insulating medium.

If the medium occupies only part of the space or if there are several media then, in addition to (2), boundary conditions also arise, namely $D_{n1} = D_{n2}$ and $E_{t1} = E_{t2}$, where the indices n and t relate to the components normal and parallel to the interface between the two media and the indices 1 and 2 indicate the media on either side of this surface.

The quantity **D** is termed the dielectric displacement. In any medium the value of ε differs from the value ε_0 applying in vacuum. The quotient

$$\frac{\varepsilon}{\varepsilon_0} = \varepsilon_r$$

is called the (relative) dielectric constant of the medium.

The electric current

A charge in motion represents an electric current. If **v** is the velocity of the charge Q then $\mathbf{I} = Q\mathbf{v}$ is the current and, in the case of an extensive charge, $\mathbf{S} = \varrho\mathbf{v}$ is the current density.

If S_n is the component of **S** perpendicular to an element $\mathrm{d}A$ of a closed area A then

$$\oint\!\!\oint S_n \, \mathrm{d}A = -\frac{\mathrm{d}}{\mathrm{d}t} \iiint \varrho \, \mathrm{d}\tau. \quad . \quad . \quad . \quad . \quad (3)$$

The measuring of charges can also be reduced to the measuring of currents (think of the ballistic galvanometer). If we proceed in a manner similar to that described above for charges, the measuring of currents can in turn be reduced to the measuring of the forces that the conductors exercise one upon the other. We shall not attempt to explain this in detail but refer to textbooks on the subject. We would only remind the reader that a current conductor determines a magnetic field characterized by a vector **B**. A part of a second conductor (length $\mathrm{d}l$) through which a current **I** is flowing and which is perpendicular to the vector **B** is subjected to a force of the magnitude

$$K = B \cdot I \cdot \mathrm{d}l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

In the absence of a coefficient this equation determines the unit for B; the direction of the force is given by Fleming's left-hand rule.

As regards the current giving rise to the field, we have it that in the vacuum the line integral $\oint B_s \mathrm{d}s$ taken along a closed curve surrounding the conductor is proportional to the current I flowing through the conductor:

$$I = \iint S_n \, \mathrm{d}A \backsim \oint B_s \, \mathrm{d}s.$$

In the same manner as for the electrotastic field, we can introduce beside the vector **B** a vector **H** which is proportional to **B** and for which the equation holds:

$$\oint H_s \, \mathrm{d}s = I, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

whilst **H** and **B** bear the relation:

$$\mathbf{H} = \frac{1}{\mu} \mathbf{B}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The advantage of this method of reasoning is again that we include at the same time the phenomena obtained when the vacuum is wholly or partly replaced by a paramagnetic or diamagnetic medium. In such a medium μ has a constant value, which differs however from the value μ_0 applying in the vacuum. The quotient

$$\frac{\mu}{\mu_0} = \mu_r.$$

is termed the relative permeability. For paramagnetic substances it is greater than unity and for diamagnetic substances less than unity.

Where there is more than one medium, separated by an interface, or where the vacuum is partly filled, there occur on the interface certain boundary conditions analogous to those for **D** and **E**. Besides the paramagnetic and diamagnetic media, ferro-magnetic media also occur. The phenomena occurring in these ferro-magnetic substances can again be described by the equations given above but then the relation between **B** and **H** is more complicated. In the simplest case (absence of hysteresis) $B = \mathrm{f}(H)$, where $\mathrm{f}(H)$ represents a function characteristic for the medium.

Variable fields [3])

So far we have been considering the case of charges in a state of rest and stationary currents. In the case of currents changing with the time t the law of induction applies, which says that: if $\iint B_n \mathrm{d}A = \Phi$ represents the surface integral of **B** across a surface bounded by a closed curve (Φ is called the magnetic flux) then there acts along the curve an electric force E_s according to the equation:

$$\oint E_s \mathrm{d}s = -\frac{\mathrm{d}}{\mathrm{d}t} \Phi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

---

[3]) See also the article by W. Opechowski, Philips Techn. Rev. **10**, 14-26, 1948 (No. 1).

E and Φ are related by the "corkscrew rule".

In equation (7) there is no new coefficient and the equation sign applies as in (4). This is closely related to the law of the conservation of energy, which connects (4) and (7).

As Maxwell first demonstrated, in the case of a surface partly bounded by a closed curve there are values of $\oint H_s \mathrm{d}s$ differing from zero not only when a charge is flowing through the surface but also when the electric flux $\Psi = \iint D_n \mathrm{d}A$ bounded by the said curve changes with time, in which latter case

$$\oint H_s \mathrm{d}s = \frac{\mathrm{d}}{\mathrm{d}t}\Psi. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

As a rule, when moreover a charge passes through this area then

$$\oint H_s \, \mathrm{d}s = I + \frac{\mathrm{d}}{\mathrm{d}t}\Psi. \quad . \quad . \quad . \quad . \quad (8a)$$

From eq. (7) we can deduce that for a closed area

$$\oint\!\!\oint B_n \, \mathrm{d}A = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

whilst from eq. (8a) it follows that

$$\oint\!\!\oint (S_n + \dot{D}_n) \, \mathrm{d}A = 0, \quad . \quad . \quad . \quad (3a)$$

which is agreement with (3).

Maxwell's equations

The equations (1), (9), (7) and (8*a*) are the well-known Maxwell equations, which together with (2) and (6) determine the properties of the electro-magnetic field. Side by side with these we have the physically less fundamental but in practice important Ohm's law, according to which, in a large number of cases, there is a proportionality between the current density S and the electric field vector **E**:

$$\mathrm{S} = \gamma \, \mathbf{E}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where γ represents a material constant, the specific conduction.

The Maxwell equations are generally given in a differential form instead of in the integral form; they then read:

$$\begin{aligned} \operatorname{div} \mathbf{D} &= \varrho, \\ \operatorname{div} \mathbf{B} &= 0, \\ \operatorname{curl} \mathbf{E} &= -\dot{\mathbf{B}}, \\ \operatorname{curl} \mathbf{H} &= \dot{\mathbf{D}} + \mathrm{S}, \end{aligned}$$

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}, \quad \mathrm{S} = \gamma \mathbf{E}.$$

One of the most important consequences of these equations is that an electromagnetic plane wave is propagated with the phase velocity

$$v = \frac{1}{\sqrt{\varepsilon \mu}}.$$

In vacuum this is the velocity of light c, so that

$$\varepsilon_0 \mu_0 c^2 = 1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The Giorgi units

As we have seen above, the units of all quantities occurring in electro-magnetism are fixed as soon as a choice has been made for the mechanical units of length, mass and time and, further, a unit has been fixed for the charge or the current intensity.

In 1901 G. Giorgi 4) put forward a number of proposals in connection with these units:

1) In the the place of the centimeter, gram-mass and second (the basic units of the c.g.s. system, to which belong the dyne as unit of force and the erg as the unit of energy) Giorgi takes the meter (m), the kilogram-mass (kg) and the second (sec.). The unit of force in this system is the force which induces in 1 kg mass an acceleration of 1 meter/sec./sec. This unit of force is called the newton (N):

$$1 \text{ newton} = 10^5 \text{ dyne}.$$

The mechanical unit of work is the work performed by the force of one newton when the point of application moves 1 meter in the direction of the force. This unit is called the newton-meter = 1 joule (J);

$$1 \text{ joule} = 1 \text{ newton-meter (N·m)} = 10^7 \text{ erg}.$$

Thus the mechanical unit of work becomes equal to the electrical unit:

$$1 \text{ joule} = 1 \text{ watt·second} = 1 \text{ VAsec}, \quad . \quad (12)$$

provided one starts from the so-called absolute volt and ampere.

2) As unit of current intensity the ampere (A) is chosen. As we have just seen, it is best to understand by this the absolute ampere.

This may be defined, for instance, as the current required to flow through two infinitely long straight conductors (with negligible circular cross section) so that, given a mutual distance of one meter, the force per meter length equals $2 \cdot 10^{-7}$ newtons (resolution of the "Comité International des Poids et Mesures", to take effect as from January 1st 1948).

All other units follow from this choice. For the greater part they coincide with those of the already commonly used "practical" system of electric units, namely with the so-called absolute practical units.

What is new in this system, however, is the fact that Giorgi consistently uses the volt and ampere also in the units for the electromagnetic field quantities.

---

4) G. Giorgi, Unità razionali di electtromagnetismo, Atti dell' Assoc. electtr. Ital. **5**, 402-418, 1901.

The quantity E is measured in volts/meter, the quantity D in coulombs/m² ($\Psi = \iint D_n \mathrm{d}A$ in coulombs) and B in newtons/ampere·m, or, in other words, in volts·seconds/m².

The last manner of writing the quantity B reminds us of another way of determining B, *viz.* by applying the law of induction. The surface intregal Φ is measured in units of volt·second, also called the weber (Wb). For H the unit is ampere/m. The units for ε and μ are A·sec/V·m = farad/m and V·sec/A·m = henry/m respectively.

Rationalization

When using c.g.s. units one is accustomed to find factors of 4π in the formulae analogous to (1) and (5). In the "mixed system of **Gauss**" the formulae (1) and (5) generally read:

$$\oint\!\!\oint D_n' \,\mathrm{d}A = 4\pi Q,$$

$$\int H_s' \,\mathrm{d}s = 4\pi I,$$

where Q and D' are expressed in electrostatic units and I and H' in electromagnetic units. Here the field quantities have been accented to distinguish them from the quantities D and H meant in this article, which are determined by (1) and (5). The omission of the factor 4π, which in the case of the quantities D and H leads to a natural interpretation such as we have already applied above, results in certain differences between the old and the new form of known formulae. For instance **Coulomb's** electrostatic law now reads:

$$K = \frac{1}{\varepsilon}\,\frac{Q_1 Q_2}{4\pi\, r^2}, \quad \ldots\ldots \quad (13)$$

and the formula for the force which two infinitely long conductors exercise upon each other over a distance l reads:

$$K = \mu\,\frac{I_1\, I_2\, l}{2\,\pi\, r}. \quad \ldots\ldots \quad (14)$$

On the other hand the formula for the capacitance C of a plane capacitor with area A and distance between the plates s becomes:

$$C = \varepsilon\,\frac{A}{s},$$

and that of the self-inductance L of a long coil of n windings with area A for a coil length s becomes

$$L = n^2\,\mu\,\frac{A}{s}.$$

It is therefore seen that the number π does not occur in formulae representing "homogeneous" cases but does appear in formulae relating to cases with spherical or cylindrical symmetry. This so-called **rationalization** (it is rational in that π occurs in the latter cases and not in the former ones) is an advantage of the manner of writing formulae (1) and (5) proposed by **Giorgi** and of the definitions for D and H embodied therein.

From formula (14) it follows, further, that since in vacuum $K = 2\cdot10^{-7}$ N, when $r = 1$ m, $l = 1$ m, $I_1 = I_2 = 1$ A,

$$\mu_0 = 4\pi/10^7 \text{ V·sec/A·m} \quad \ldots\ldots \quad (15)$$

Taken together, (15) and (12) can also be construed as a way of defining the "absolute" electrical units.

Since $\varepsilon_0\mu_0 c^2 = 1$, with $c = 2.99776\cdot10^8$m/sec it follows that

$$\varepsilon_0 = \frac{10^7}{4\pi c^2} \approx 8.855\cdot10^{-12} \text{ A·sec/V·m} \quad \ldots \quad (16)$$

A remark is to be made in respect to the **Coulomb** magnetic law. As is known, a long coil acts as a rod magnet with "poles" at the extremities. From the north pole there is a flux $\Phi = \iint B_n \mathrm{d}A$ (taking the integral across the cross section of the coil). This flux can be taken as a measure of the pole strength, but with a given current this varies with the medium. Independent of the medium, however, is the quantity $\iint H_n \mathrm{d}A = P$. If Φ is chosen as pole strength then the force between two poles at a distance r is:

$$K = \frac{1}{\mu}\,\frac{\Phi_1\Phi_2}{4\pi r^2} \quad \ldots\ldots\ldots\ldots \quad (17)$$

If P is taken as the pole strength then

$$K = \mu\,\frac{P_1 P_2}{4\pi r^2}. \quad \ldots\ldots\ldots\ldots \quad (18)$$

Both these formulae give expression to **Coulomb's** law of magnetism. Formula (17) is, formally, analogous to (13) whilst (18) is analogous to (14). The case is more or less academic, since actually free magnetism does not exist [5]).

Historical notes

The c.g.s. systems so far usually employed go back to the two laws of **Coulomb**:

$$K_{\mathrm{el}} = a\,\frac{Q_1 Q_2}{r^2} \quad \ldots\ldots \quad (19)$$

and

$$K_{\mathrm{magn}} = \beta\,\frac{P_1\, P_2}{r^2}, \quad \ldots\ldots \quad (20)$$

where P_1 and P_2 represent the magnet pole strength, which laws have been formulated in analogy with **Newton's** law of attraction (about 1680).

After **Oersted** (1819) had determined how electrical currents act upon the poles of a magnet [6]) it was particularly Ampère [7]) (Sept.-Dec. 1820) who undertook a thorough investigation of these forces, as also the forces between conductors. With the aid of what is understood as the "current element" one can formulate Ampère's results in the same way as the Coulomb laws. The most familiar formula is that of Laplace [8]), which reads:

---

[5]) See also A. **Sommerfeld**, Über die elektromagnetischen Einheiten, Z. techn. Phys. **16**, 420-424, 1935, and "Verhandelingen aangeboden aan Prof. Dr. P. **Zeeman**", M. Nijhoff, The Hague, 1935, pp. 157-165.

[6]) **Oersted** spoke of the "electrical conflict". The title of his paper of 1820 read: "Experimenta circa effectum conflictus electrici in acum magneticum". He imagined that when a **Volta** battery is circuited with a wire the "two electricities" (+ and —) combine and that this turbulent process causes the magnetic action.

$$K = \delta \frac{I\,ds \cdot P}{r^2} \sin \varphi, \quad . \; . \; . \; . \; . \; . \; . \qquad (21)$$

indicating the force exercised by a current element Ids upon a magnet pole P at a distance r (φ is the angle between the directions of r and ds, K being perpendicular to those directions, whilst δ is a constant).

It was Ampère, too, who first introduced the concepts of current and tension as we know them to-day. The concept of resistance and the laws indicating the relation between current and voltage when applied to conductors are due to Ohm (1829).

From the formulae (20) and (21) Gauss developed (about 1830) the so-called electromagnetic system of units, following upon which Weber (about 1850) derived from (19) the electrostatic system. These systems were not officially adopted as c.g.s. systems until 1873. (Gauss worked at first with the millimeter as length unit and the milligram as mass unit).

In 1831 Faraday discovered the law of induction (7) and in 1864 Maxwell revealed its analogue (8). The introduction of the field concept began with Faraday and Maxwell. As is known, Maxwell was the first to prove that the relation between the electromagnetic and the electrostatic charge units is equal to the velocity of light (in centimeters per second).

Helmholtz (1882) and Hertz (about 1890) employed a mixed system of units where electrical units were measured in terms of e.s.u. and magnetic units in e.m.u. (generally called the "Gauss mixed system"). The three c.g.s. systems are characterized by the place occupied by the factors c ($\approx 3 \cdot 10^{10}$) and c^2 in the formulae.

The "practical" units

None of the above-mentioned systems of units was suitable for practical use in the developments taking place in electrotechnics. Through the activities of the British Association for the Advancement of Science (1867) there came to be adopted the now commonly used practical units of the ohm, volt and farad, to which were added by the "Congrès International des Electriciens" (1881) [9]) the coulomb, ampere and henry (originally sec·ohm = ohm·sec).

These units are equal to the electromagnetic units but for factors of powers of 10: 1 ohm = 10^9 e.m.u., 1 volt = 10^8 e.m.u., 1 farad = 10^{-9} e.m.u., 1 coulomb and 1 ampere = 1/10 e.m.u., 1 henry = 10^9 e.m.u. In 1935 the I.E.C. ("International Electrotechnical Committee") decided to apply Giorgi's M-K-S system in electrotechnics.

International units

Together with the adoption of the practical units there arose a need to define them by standards or standardized specifications. It would lead us too far to go into this at any length here and we shall only briefly relate how the so-called international ampere and ohm have so far come to be established:

1 int. amp. = the current which when passed through a specified silver voltameter will deposit silver at a rate of 1.11800 mg/sec.;

1 int. ohm = the resistance of a column of mercury of uniform cross-sectional area, 1.06300 m in length and 14.4521 g in mass at 0 °C.

From these two data there follows the international volt, which, moreover, has also been fixed by taking the e.m.f. of the Weston cell at 10°C as equal to 1.01830 V.

As was to be expected, upon further investigation these units were found to be practically but not precisely in agreement with the "absolute" or "theoretical" units. Various government laboratories have since then determined the relation of the international to the absolute practical units. The average equations of six laboratories (Germany, England, France, Japan, U.S.A., U.S.S.R.) are as follows [10]):

1 int. ohm = 1.00049 absolute ohm,
1 int. amp = 0.99985 absolute ampere,
1 int. volt = 1.00034 absolute volt,
1 int. watt = 1.00019 absolute watt (10^7 erg/sec).

Although the deviations are so small as to be of no importance for many practical measurements, where measurements have to be precise it is indeed of importance to know exactly what units have been used. In particular, the double definition of the international volt is apt to lead to mistakes [11]).

---

[7]) See e.g. G. A. Boutry, Ampère ou la Pureté, Rev. trim. Canad. **33** 257-274, 1947 (No. 131).

[8]) Commonly known as the Biot-Savart law (1829), although these authors were only considering the special case of an infinitely long straight conductor.

[9]) See e.g. G. Giorgi, La métrologie classique et les systèmes d'unités qui en dérivent, Examen critique. Rev. Gén. Electricité **40**, 457-467, 1936.

[10]) See e.g. Mesures **11**, 379-380, 1946.

Mixed use of practical and c.g.s. units

Whilst the practical units came into general use for research as well as for engineering — largely as a result of the availability of standard instruments calibrated in these units, such as resistance boxes and meters — c.g.s. units continued to be used for describing electrical and magnetic fields. For this purpose powers of 10 are introduced in the formulae as coefficients, as for instance in the formula

$$V = -10^{-8}\frac{\mathrm{d}\Phi}{\mathrm{d}t},$$

indicating the relation between the electromotive force V induced in a winding of a transformer and the magnetic flux Φ, where V is expressed in volts and Φ in gauss·cm^2 (= maxwell). We have already seen how the proposals made by Giorgi have led to these factors also disappearing from the formulae.

Realization of the absolute ampere, ohm and volt

In 1910 at a conference held in Washington agreement was reached between the laboratories of the U.S.A., England, France, Germany, Japan and Russia in regard to the standards of resistance and e.m.f. In 1930 however the standards of the various countries were found to show intolerable discrepancies. The "Comité International des Poids et Mesures" has now decided to return as from January 1st 1948 to the absolute units, on the ground that modern measuring technique has been sufficiently developed to be able to realize these units at any desired moment. Several different methods of doing this are in use [12]).

One method for determining the ohm as an absolute measure, for instance, is to compare in a bridge circuit the self-inductance of a coil of known dimensions with the capacitance of an auxiliary capacitor, which capacitance is then compared with a resistance in a second measuring bridge. The absolute ampere is determined by the force acting between two conductors in accordance with the definition given above. Of course one does not use infinitely long straight conductors, but circular coils of exactly known dimensions and with a precisely known mutual position. The force is measured with a current balance. Since the force depends not only on the masses of the weights but also on the acceleration due to gravity at the place where the measurement is taken, it is necessary to measure the latter very accurately. When the absolute ampere and ohm are known, then the absolute volt is the voltage induced by a current of one ampere at the extremities of a resistance of 1 ohm.

The international units now play at most the part of secondary standards. That is why we have given above the relation between absolute and international units, as fixed for 1948. In a subsequent article in this journal some practical consequences of the use of the rationalized Giorgi system (with absolute volt and ampere) will be explained further.

---

[11]) See e.g. U. Stille, Die Umrechnungsfaktoren von internationale auf absolute elektrische Einheiten, Z. Phys. **121**, 24-53, 1943.

[12]) See e.g. L. M. Briggs, Rev. mod. Phys. **11**, 111-120, 1939.

CONICAL DISC SPRINGS

by J. A. HARINGX.

621-272.4

Relatively little use is made of the disc spring as resilient element in constructions. In respect to the amount of energy that can be absorbed a disc spring can certainly lay no claim to any fundamental superiority over the helical spring, but in certain cases it does make the construction simpler and more efficient. What these cases are will be discussed in this article, following upon a brief explanation of the properties of disc springs, some of which are rather remarkable. The designing of disc springs most suitable for a certain purpose is explained in more detail for two particular cases. It appears that the most favourable shape of the disc spring (for a given load and maximum permissible stress) can be very easily determined with the aid of appropriate graphs.

Among the various kinds of springs applied in the technique the conical disc springs (often called Belleville washers) are relatively little known. In recent times, however, they have been receiving more attention because they often help to simplify constructions.

A conical disc spring — in the following pages the adjective "conical" will mostly be omitted — consists of what we might call the "collar" of a truncated cone the vertex angle of which is approximately 180° (see *fig. 1*). In the limit case where this vertex angle is exactly 180° we have a round flat plate with a circular aperture at the centre. Although the disc spring therefore bears a certain resemblance to such a flat plate, the results of the theory of bent plates may not be used directly for disc springs, not even as an approximative solution. This is due to the entirely different nature of the stresses occurring. For instance, if a disc spring supported underneath is loaded with an axial force uniformly distributed along the upper rim it undergoes an axial compression and takes the shape as indicated by the full line in *fig. 2*. The outer edge is forced slightly outwards and stretched, whilst the inner edge is shortened. As a consequence, in addition to bending stresses there also arise radial and tangential stresses uniformly distributed over the thickness, which stresses do not occur in bent plates, at least not at small deflections.

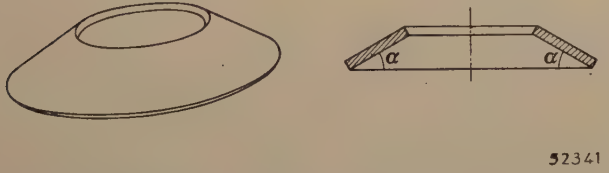

Fig. 1. Diagrammatic representation of a conical disc spring in the non-loaded state. On the left a view in perspective, on the right the meridian cross section. The angle α is actually much smaller than that indicated here, seldom exceeding 6°.

In this connection disc springs show some typical properties which will be discussed in this article. Then we shall consider how the disc spring should be shaped for a given load and maximum permissible stress in two particular cases, and finally we shall give some examples of the application of disc springs.

Properties of the disc spring

General properties

Looking at fig. 2 we see that the top and bottom edges of the disc spring lie in two parallel planes, not only in the original state but also in the loaded state, on account of the axial symmetry. If the spring is placed between two parallel "compression plates" the compressive force will therefore always be transmitted in an axial direction. Moreover, the shallow height allows of an element with great resilience being used in a small space. These two properties are highly important and in many cases make the application of the disc spring exceptionally attractive.

Furthermore, it is possible to combine several of these disc springs (say n in number) in the manner illustrated in *fig. 3a*. With this arrangement, given the same load, the compression is n times as great as that for one single disc. If, on the other hand, n

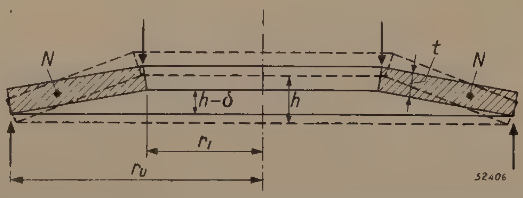

Fig. 2. The meridian cross section of a loaded disc spring resulting from the meridian cross section of the non-loaded spring (in broken lines) after rotation around the points N. Both in the loaded and in the non-loaded state the top and bottom edges lie in two parallel planes. The inner and outer radii of the spring are indicated by r_i and r_u; h and t are respectively the height and thickness of the spring; δ is the compression.

springs are stacked in the manner indicated in fig. 3*b* a load *n* times as great is required to bring about the same compression. Of course both these methods could be used in combination one with the other, but it has to be pointed out that the system of stacking into what might be called a packet (fig. 3*b*) always gives rise to friction and wear, which is often troublesome.

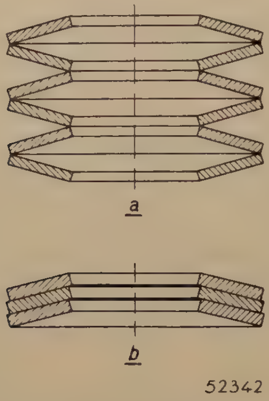

Fig. 3. *a*) A resilient column built up with disc springs. *b*) A resilient packet built up with disc springs.

The amount of energy that can be absorbed per unit of volume of the material is about a factor 3 less in the case of the disc spring than in the case of the helical spring. Thus, to reach the same effect, about three times as much material is required when using a disc spring. For the same amount of material the disc spring does, it is true, take up less space than the helical spring, but not so much less as to compensate this factor 3. When it is a matter of accommodating the largest possible amount of energy in a given space the disc spring is at a disadvantage by a factor of about 2, so that from this point of view there is no purpose in using a column of disc springs in the place of a helical spring. However, there are other reasons why a disc spring may after all be preferable as compared with the helical spring. In certain cases it is useful, for instance, to be able to adjust easily the rigidity of the resilient element (i.e. the load per unit of compression); whereas the rigidity of a helical spring already made cannot be altered, this is indeed possible with a column built up out of several separate disc springs. A further important point is that disc springs can be made in any engineering works, whereas the coiling of heavy helical springs requires some experience.

Relation between load and compression

Apart from the general characteristics of disc springs mentioned above, which are generally decisive for their application, in certain cases it may be advantageous to make use of the peculiar shape of the curves representing the relation between load and compression. We shall prove this with the aid of the theoretical formula expressing the said relation, viz:

$$\frac{P}{P_h} = 1 + \left[1 - \tfrac{1}{2}\left(\frac{h}{t}\right)^2\right]\left(\frac{\delta}{h} - 1\right) + \tfrac{1}{2}\left(\frac{h}{t}\right)^2\left(\frac{\delta}{h} - 1\right)^3. \quad (1)$$

The meaning of the symbols *h*, *t*, *δ* occurring in this formula can be found in the subscript to fig. 2: *h* and *t* are respectively the height and thickness of the disc spring; *δ* is the compression. Further, *P* is the axial load uniformly distributed over the edges of the disc spring, whilst P_h represents the force required to flatten the disc spring ($\delta = h$). A calculation shows that

$$P_h = \frac{Eht^3}{Mr_u^2}, \quad \dots\dots \quad (2)$$

where *E* is the modulus of elasticity, *M* a certain function of r_u/r_i, and r_u and r_i are respectively the external and the internal radius of the disc spring (see fig. 2).

From equation (1) it is seen that the relation between load and compression depends only upon the thickness of the spring, taking P_h as load unit and *h* as length unit.

Fig. 4 shows the relation between P/P_h and δ/h according to eq. (1) for various values of h/t. The curves appear to fall into two groups according as $h/t > \sqrt{2}$ or $h/t < \sqrt{2}$.

Where $h/t > \sqrt{2}$, that is to say for disc springs that are relatively thin, the curves show a maximum and a minimum. When a disc spring is compressed continuously, for instance by placing it between two parallel plates drawn closer together by turning a screw, the compressive force exercised by the disc spring upon the compression plates does not by any means change at a uniform rate, but in a certain range (between the maximum and the minimum) it decreases with increasing compression. When the ratio h/t is greater than $2\sqrt{2}$, the (P, δ)-curve intersects the abscissa, so that corresponding to a compression $\delta > \delta_1$ (see in fig. 4 the curve for $h/t = 3$) we find a negative compressive force. Since in this case a disc spring is incapable of exercising a negative compressive force (thus a tensile force) the spring loses contact with the compression plates and, being now unloaded, changes in shape until the stable state of equilibrium $\delta = \delta_2$ is reached; in other words the disc spring suddenly gives [1]).

We now consider δ as a function of P instead of P as a function of δ, that is to say, instead of bringing the two compression plates closer together by means of a screw, we bring a continuously increasing force to bear upon the spring, for instance with the aid of weights. It will be clear that the disc spring will now give as soon as the force reaches the maximum in the (P, δ)-curve and the compression $\delta = \delta_1'$ will jump to the value $\delta = \delta_2'$ (see in fig. 4 the curve for $h/t = 2$).

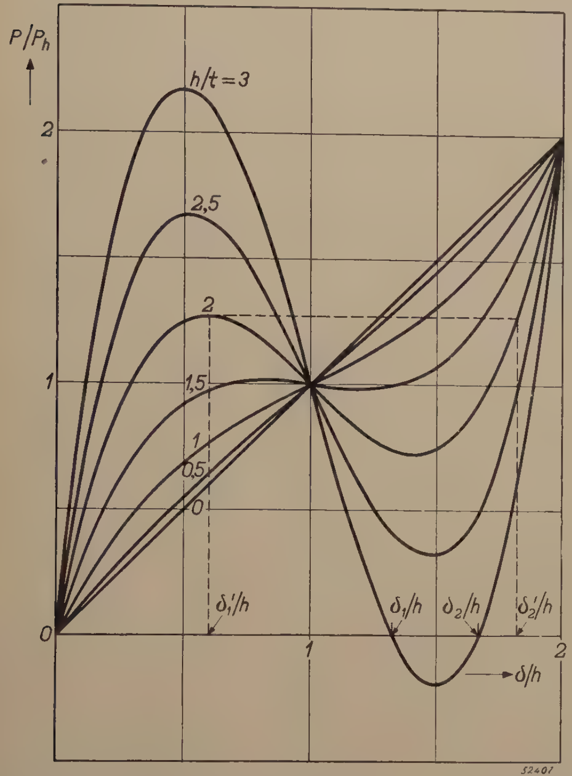

Fig. 4. The relation between the load P and the compression δ of a disc spring for different values of the ratio of the height h to the thickness t. The length unit is the height h of the disc spring and the load unit is the value P_h of the load at the flattened state of the spring.

Usually it is desired to avoid these phenomena of "giving". A construction must therefore be designed in such a way that the compressions actually occurring are always less than the compression corresponding to the maximum of the (P, δ)-curve.

Where h/t is less than $\sqrt{2}$ on the other hand the compressive force always increases at a uniform rate with the compression. The greater the thickness of the disc spring, the more the (P, δ)-curve approaches a straight line.

For $h/t = \sqrt{2}$, the curve representing the compressions round about $\delta = h$ runs practically horizontal. It is this peculiar property of the disc spring that is of great practical importance in certain constructions, in particular in those cases where the load is required to be independent of the compression. An example of such an application of the disc spring will be given presently.

Relation between the maximum stresses and the compression

An important property of any spring in practice is the relation between the maximum stresses occurring in it and the load or corresponding compression. The maximum stress may never exceed a permissible limit, which is to be regarded as a given material constant. In the case of a disc spring the maximum (compressive) stresses σ occur along the top rim on the inside. For σ the following formula holds

$$\sigma = \frac{E\delta h}{M{r_u}^2}\left[C_1\left(1-\frac{\delta}{2h}\right)+C_2\frac{t}{h}\right], \quad \left(\frac{\delta}{h}<2\right), \ldots \quad (3)$$

where C_1 and C_2 are certain functions of r_u/r_i given by the theory; the other terms have already been defined.

Theoretical fundamentals of equations (1) to (3)

We shall now say a few words about the manner in which the behaviour of the disc spring represented by the equations (1) to (3) can be deduced. Since the problem of the disc spring has never yet been dealt with in an exact theoretical manner, we can only work upon the approximative calculation of Almen and Laszlo [2]). These authors introduce the supposition — which is after all only obvious — that the rectangular cross section of the spring lying in the meridian plane does not change in shape when loaded, but only turns about a "neutral" point, indicated in fig. 2 by the point N. This greatly simplifies the determination of the tangential elongations ε_t. Applying this conception consistently, they put the radial elongations ε_r as zero and calculate the tangential stresses σ_t with the aid of the equation $\sigma_t = m^2E\varepsilon_t/(m^2\text{-}1)$, where m represents Poisson's ratio $= 10/3$. The radial stresses then amount to $\sigma_r = \sigma_t/m = 0.3\,\sigma_t$. The approximation theory given by Almen and Laszlo can be partly checked with the flat "conical" disc spring ($h = 0$), the behaviour of which can be calculated with the conventional theory of the bending of flat plates. The shape of the flat disc spring (for instance in the case where $r_u = 2r_i$) in the loaded state does indeed deviate only very little from a cone, so that one can safely calculate the tangential elongations ε_t as those obtaining for the conical shape. The small deviation from the conical shape is due to the fact that there are practically no radial bending stresses: the outer or inner edges of the disc spring are free of load in a radial direction, and here the radial stresses are in fact exactly zero. Towards the centre between the inner and outer edges they do, it is true, increase, but in the case that $r_u = 2r_i$ they reach at most a value of 0.065 σ_t and are therefore much smaller than the stresses 0.3 σ_t which would be present according to the calculation of Almen and Laszlo.. We may therefore expect a better approximation of the problem of the disc spring when we put instead of the radial elongations the radial stresses at zero, thus calculating the tangential stresses with the aid of the equation $\sigma_t = E\varepsilon_t$. As a result of this correction the factor $1\text{-}\sigma^2 = (m^2\text{-}1)/m^2$ is eliminated from the denominators in the formulae of Almen and Laszlo. At the same time the differences between their calculation and the experimental result are likewise partly eliminated.

---

[1]) We have a familiar example of this phenomenon in the case of certain containers closed by a metal lid that is slightly convex. When more than a certain limit pressure is applied to the centre of the lid this gives inward and the serrated rim is forced out.

[2]) J. O. Almen and A. Laszlo, The Uniform Section Disk Spring, Trans. Amer. Soc. Mech. Engrs. 58, 305-314, 1936.

Except for the factor just referred to, eqs (2) and (3) given above are identical with those of Almen and Laszlo.

Design of disc springs

It is not the intention to deal here with the technical problem of the designing of a disc spring, or of columns of disc springs, in general terms. All we intend to do is to show briefly, with reference to two special cases, how the shape of a disc spring can best be chosen with the aid of suitable graphs.

First case

In the construction of a press in one of the Philips factories it was found useful to employ columns of disc springs built up in the manner indicated in fig. 3*a*. In order to minimize the number of springs in a column like this one has to use disc springs with a certain optimum value of the ratio of the external diameter to the internal diameter. For the greatest possible compression under a given load and given maximum stress this ratio should preferably be chosen equal to 2, as may be determined by a calculation applying for the case of a flat disc spring. Thus we put once for all $r_u/r_i = 2$, when the functions M, C_1 and C_2 occurring in equations (2) and (3) have respectively the values 0.69, 1.225 and 1.58.

Furthermore for constructional reasons disc springs which under the maximum load pass through the flat state cannot be used. If, therefore, we fix the minimum slope of the generatrix of the disc spring at say 0.04 we get the condition:

$$h - \delta = 0.04\,(r_u - r_i),$$

that is to say, since $r_u = 2r_i$,

$$\frac{h}{\delta} = 1 + 0.02\,\frac{r_u}{\delta} \quad . \ . \ . \ . \ . \quad (4)$$

Moreover, we have to take care that the maximum stresses σ in the spring, calculated according to eq. (3), never exceed a certain permissible limit.

To give a disc spring the greatest possible elasticity its material must allow of very great stresses. What the maximum permissible stresses are is a matter of experience. For plain carbon steel, for instance, σ may amount to 120-140 kg/mm², whilst in the case of quenched and tempered chromium-nickel steel stresses may occur as high as 180-200 kg/mm² (with thin disc springs even up to 220 kg/mm²). These are very high values and it is therefore inevitable that under these conditions local plastic distortions will arise, so that already under the first load the disc spring "settles" a little, thus reducing the height h. This must therefore always be taken into account when designing a disc spring. Should the load not be static, as tacitly understood so far, but dynamic then a repetition of such plastic deformations will in the long run prove fatal. Therefore in this case we must base our calculations on the fatigue strength of the material, which means that with a good kind of spring steel a stress variation of at most 100 kg/mm² is permissible.

Taking these requirements into consideration, not much freedom is left in the choice of the most suitable shape for a disc spring, for we have to work upon four equations (1)-(4) between seven quantities, P, P_h, δ, σ, r_u, h, t, which means that we can only freely dispose of three quantities. Now in the practical problem the load P and the maximum stress σ are given, so that really we are only free to vary one quantity, and for this it is well to choose r_u. By a simple conversion of the equations (1)-(4) we obtain two relations of the following form:

$$\frac{\delta}{r_u} = \Psi\left(\frac{P}{r_u^2}, \sigma\right) \quad . \ . \ . \ . \ . \quad (5)$$

and

$$\frac{t}{r_u} = \Phi\left(\frac{P}{r_u^2}, \sigma\right) \quad . \ . \ . \ . \ . \quad (6)$$

We shall not write down the rather complicated, though elementary, functions Φ and Ψ but give them in the form of graphs in *fig. 5*.

With P, σ and r_u given, we now find from equation (5) or from the right-hand graph of fig. 5 the value of δ/r_u, and from eq. (6) or the left-hand graph of fig. 5 the value of t/r_u. Only h is still unknown, and this we find from eq. (4).

The curves in the right-hand graph of fig. 5 are limited upward by a broken line defined by the condition $t/r_u < 0.25$. This condition has to be fixed because otherwise the thickness of the disc spring would exceed the practical limit. The bottom limit in the graph has been introduced in order to avoid too strong a deviation from the straight (P, δ) line and also as a safeguard against the troublesome "give" phenomena (cf. fig. 4). The upper and the lower broken lines for the bottom limit correspond respectively to the condition $h/t < 1.0$ and $h/t < 1.2$.

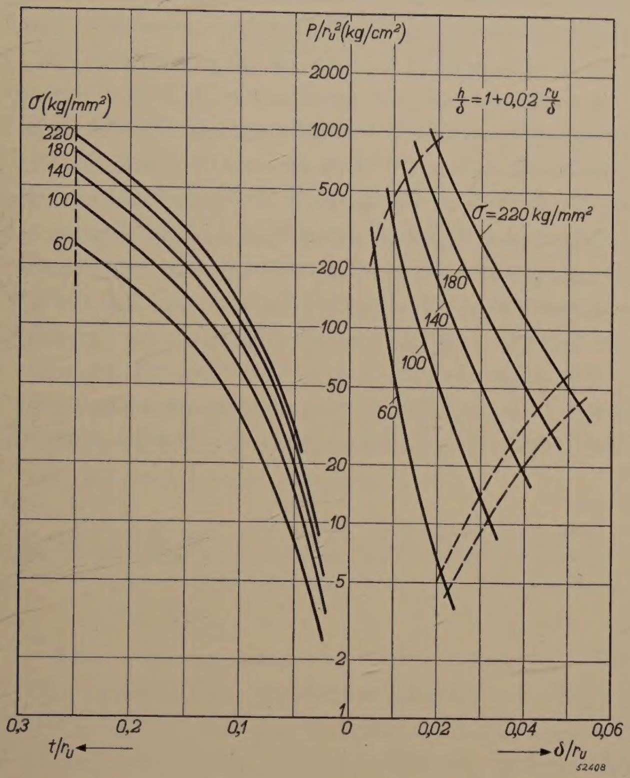

Fig. 5. Graph showing the compression δ and the thickness t of the spring for a given load P, maximum permissible stress σ and outer radius r_u ($= 2r_i$) of the disc spring. The height h of the spring is calculated with the aid of the formula given (eq. (4)). This graph holds for the case where a disc spring has to be designed which allows of the greatest possible compression under a given load.

Tests made with disc springs calculated in the manner described above showed that the actual compression was smaller than calculated. Upon further consideration a very acceptable explanation of this was found. Disc springs stacked in a column are apt to slide one over the other on account of the sharp edges, and in order to avoid this the top edge and the bottom outer edge were slightly ground to provide a flat seating (see *fig. 6*). The compression being rather great, the slope of the cone becomes noticeably smaller and the flat-ground edges do not remain flat but become slightly conical, so that the original contact planes of two adjacent disc springs begin to show some clearance. As a consequence the load is transmitted at circles with radii r_i' and r_u' respectively instead of at those of radii r_i and r_u respectively. Obviously $r_i' > r_i$ and $r_u' < r_u$.

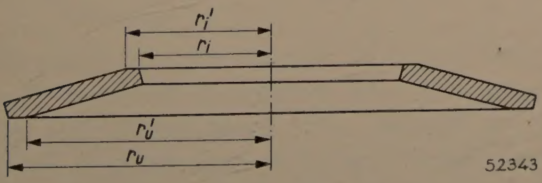

Fig. 6. When stacked up in a column the disc springs may easily slide over each other. To avoid this the top and bottom edges are slightly ground off to a flat plane as sketched here.

Taking this fact into consideration we can still use eqs (5) and (6) as well as the corresponding graphs in fig. 5, with hardly any alteration. As a further consideration will show, we only have to introduce in place of the actual load P a fictitious load $P' = kP$ where

$$k = \frac{r_u' - r_i'}{r_u - r_i}.$$

However, the value then obtained for δ', from the modified eq. (5),

$$\frac{\delta'}{r_u} = \mathrm{f}\left(\frac{P'}{r_u^2}, \sigma\right),$$

is not the actual compression δ. In order to arrive at δ, the value δ' has to be multiplied by the factor k:

$$\delta = k\delta' \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (8)$$

The compression values calculated in this manner agree well with the measured results.

Second case

As already stated, a typical property of the disc spring is the fact that when it is suitably dimensioned its compression can be varied between wide limits without hardly any change in load. If this property is to be utilised there must be a certain relation between the height and the thickness of the spring, namely $h/t \approx \sqrt{2} = 1.41$. For instance for

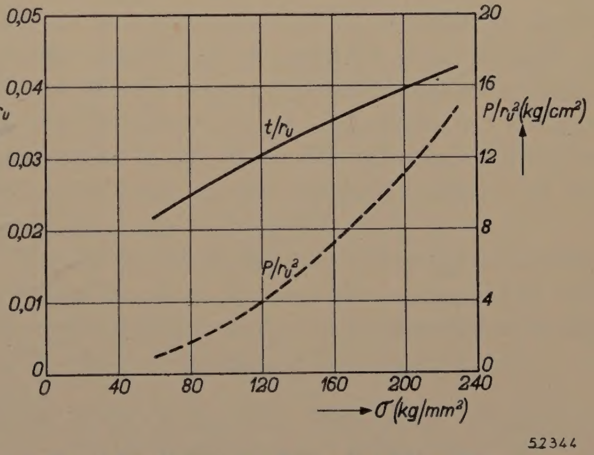

Fig. 7. Graph for designing a disc spring where the compression is required to be variable within wide limits under a constant load. In this case the ratio of the height h to the thickness t is 1.5. For a given load P and maximum permissible stress δ this graph gives the values of the outer radius r_u ($= 2r_i$), the thickness t and the height h.

$h/t = 1.5$, P is found to be constant within the region $0.6\,h < \delta' < 1.4\,h$. This constant value of P depends only upon σ and r_u. In *fig. 7*, P/r_u^2 is plotted as a function of σ for $h/t = 1.5$ and $r_u = 2r_i$. When the load is given one can therefore find immediately from this graph what value has to be chosen for r_u (and r_i). Fig. 7 also gives the curve representing t/r_u as a function of σ, from which can be derived the required thickness and height of the spring for a given r_u.

Possibilities of application

The example already mentioned where a column of disc springs was required for a press lies in a domain where also a helical spring can be used. As stated when dealing with the properties of disc springs, in this case the disc spring does not offer any fundamental advantages over the helical spring and only secondary circumstances are decisive.

The specific field of application for the disc spring is rather to be sought in cases where the helical spring cannot be considered. For instance, advantage can be taken of the shallowness and excellent centering of the compressive force in the case of friction couplings, for fixing securely the parts of a motor commutator and wherever an elastic construction is desired to take up large impulses [3]).

An example of the application of a disc spring for a compressive force independent of the compression is given by Almen and Laszlo in the article quoted in footnote [2]). There it concerns the fixing of the tail-stock centre of a lathe. Owing to the generation of heat in the turning process the workpiece expands and if the centre does not "give" the workpiece gets jammed. By applying the disc spring construction this is avoided, for here the centre has axial freedom of movement without the axial force increasing to any appreciable extent.

---

[3]) Examples given by G. Ashworth, The Disk Spring or Belleville Washer. Proc. Inst. Mech. Engrs **155**, 93-100, 1946.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

R 53: **A. van der Ziel**: Method of measurement of noise ratios and noise factors (Philips Res. Rep. **2**, 321-330, 1947, No. 5).

In this paper methods are described, as applied in this laboratory, for the measurement of noise ratios of impedances and of the noise factor of receivers. Instead of standard signal generators, saturated diodes are used as "standard noise generators". The noise voltages are amplified in a linear amplifier having a relatively small bandwidth (50-100 kc/sec), and detected by a thermocouple.

R 54: **J. L. H. Jonker**: Reflections in electron tubes (Philips Res. Rep. **2**, 331-339, 1947, No. 5).

The characteristics of some electron tubes show irregularities as a result of reflected electrons. By studying the peculiarities of these reflections at low potentials, it is possible to deduce their influence on the electric field between the electrodes, and in this way to explain the irregularities.

R 55: **F. A. Kröger**: The temperature dependence of the fluorescence of tungstates and molybdates in relation to the perfection of the lattice (Philips Res. Rep. **2**, 340-348, 1947, No. 5).

The quenching temperature of the photoluminescence of tungstates and molybdates is dependent on the perfection of the crystals, imperfect crystals having a low quenching point. This property may be explained in terms of the quenching theories of Mott and Seitz, as well as in terms of the theories of Peierls, Möglich - Rompe and Frenkel.

R 56: **F. de Boer**: Some characteristics of trigonal selenium crystals obtained from the vapour phase (Philips Res. Rep. **2**, 349-351, 1947, No. 5).

Trigonal selenium crystals of a size sufficient for conduction measurements have been prepared from the vapour phase. The majority of these crystals had the form of hollow needles showing slight deviations from the ideal monocrystal structure. A small fraction, however, consisted of thick and massive twin crystals, amongst which one rather well-developed monocrystal was found. All well-developed planes were prism planes.

R 57: **F. de Boer**: On the electrical conductivity of selenium crystals (Philips Res. Rep. **2**, 352-356,1947, No. 5).

For the specific resistance of selenium monocrystals parallel to the c-axis values ranging from 2.10^4 to 5.10^4 Ω cm were observed; it is made plausible that the specific resistance of a pure monocrystal is lying near the lower end of this range. The specific resistance at right angles to the c-axis was found to be 2.10^6 Ω cm. In addition some experiments on the influence of temperature and pressure on the conductivity were carried out.

R 58: **L. J. Dijkstra**: Elastic relaxation and some other properties of the solid solutions of carbon and nitrogen in iron (Philips Res. Rep. **2**, 357-381, 1947, No. 5)

The theory of elastic relaxation in α-iron caused by carbon and nitrogen in solid solution, as given for the first time by Snoek and later elaborated by Polder, predicts a strong anisotropy for the various crystal directions. This theoretical conclusion has been confirmed in a series of experiments carried out on prepared single crystals of iron in the (100)- and (111)-directions. For carbon it was possible to check the theoretical absolute magnitude of the effect. The question of the most probable place of the dissolved particles in the iron lattice is also discussed.

Finally at 20 °C the rate of segregation taking place in the form of a carbide or nitride has been determined by measuring the decrease in magnitude of the elastic relaxation in the course of time.

R 59 **J. D. Fast**: The dissociation of nitrogen in the welding arc (Philips Res. Rep. **2**, 382-398, 1947, No. 5)

For temperatures in the range from 5000-10000 °K the dissociation of nitrogen is computed on the basis of spectroscopic data and for three different values of the dissociation energy, viz. 7·383, 8·573 and 9·764 electron-volts; in view of a recent investigation by Gaydon and Penney the highest value is probably the correct one.

At 10000 °K and a pressure of one atmosphere the degree of dissociation is 99% at least; at 5000 °K 25.1% of the nitrogen will be dissociated if the dissociation energy amounts to 7.383 eV, but only 1.63% will be dissociated if the dissociation energy is 9.764 eV. Two cruder methods of computation give results that differ only slightly from these "exact" values.

R 60: J. Haantjes and B. D. H. Tellegen: The diode as converter and as detector (Philips Res. Rep. 2, 401-419, 1947, No. 6)

The current through a diode, to which is applied, in addition to an A.C. voltage and a D.C. voltage, a small extra voltage v, can be developed into a power series of v whose coefficients are Fourier series. The magnitude of these coefficients is calculated for a diode that has a linear characteristic in the pass direction. By confining ourselves to the term linear in v, four-pole equations and equivalent circuits can be set up for the diode as converter and as detector. From these equivalent circuits various properties and quantities can be deduced. The fluctuations of the diode as converter can also be represented with the help of the equivalent circuit. Upon conversion as well as upon detection it is found favourable to give the diode a small internal resistance.

R 61: H. C. Hamaker: Radiation and heat conduction in light-scattering material (Philips Res. Rep. 2, 420-425, 1947, No. 6)

In previous papers (see **R 35, R 38** and **R 39**) only simplified problems have been considered. For a fuller discussion of many practical questions of a similar kind some extensions of the theory are required. These are indicated, though no attempts are made to work them out in full detail.

R 62: N. Warmoltz: The time-lag in starting a low-pressure arc on a mercury or gallium cathode in connection with field emission and surface deformation (Philips Res. Rep. 2, 426-441, 1947, No. 6)

For the contents of this article see Philips techn. Rev. **9**, 105-113,1947 No. 4

R 63: W. Elenbaas: The continuous spectrum of the high-pressure mercury discharge (Philips Res. Rep. 2, 442-453, 1947, No. 6)

The intensity of the continuous spectrum of the high-pressure mercury discharge has been measured between $^1/_4$ and 25 atm for inputs from 20 to 60 watts/cm and for diameters between 6 and 60 mm. The ratio of the intensity of the continuum to the intensity of the yellow lines 5770/91 is almost independent of the input and increases linearly with the mean vapour density m/d^2. The ratio at $m/d^2 = 0$ is interpreted as being due to recombination of electrons with ions, whereas the contribution proportional to m/d^2 originates from molecular radiation.

A rough energy balance is given in which the line energies and the energies radiated in the two types of continuum occur. Using the equation, the absolute value of the intensity of the continuum of the U.V.-standard is found to be more than twice as high as the value measured by Rössler.

The discrepancy between the measured and calculated gradient between 1 and 5 atm disappears on using this energy-balance equation instead of that formerly used, in which the lines and the recombination spectrum were the only radiations taken into account. The remaining discrepancy at higher pressures is probably due to the diminution of the ionization potential at these high densities. The magnitudes of these diminutions necessary to account for the measured gradients are tabulated as a function of the mean vapour density.

R 64: J. W. L. Köhler and C. G. Koops: Absolute measurement of the time constant of resistors (Philips Res. Rep. 2, 454-467, 1947, No. 6)

A new method is described for the accurate determination of the time constant of resistors in absolute measure, using as standards of reference a set of standard condensers with negligible losses as described in a previous paper (see Philips techn. Rev. **5**, 311-319, 1940)

R 65: F. L. H. M. Stumpers: On the calculation of impulse-noise transients in frequency-modulation receivers (Philips Res. Rep. 2, 468-474, 1947, No. 6)

The effect of impulse-noise transients is calculated by means of a series expansion of the phase, the general term of which contains $[A(t)]^n$ when the amplitude $A(t)$ of the disturbance is smaller than the amplitude of the signal, and $[A(t)]^{-n}$ in the inverse case. The Laplace transform is used to calculate the effect in the filters. The large effect of phase-opposition during the capture time is accounted for.